

Theoretical and Mathematical Physics

Sergio Cecotti

Introduction to String Theory



Springer

Introduction to String Theory

Theoretical and Mathematical Physics

This series, founded in 1975 and formerly entitled (until 2005) *Texts and Monographs in Physics* (TMP), publishes high-level monographs in theoretical and mathematical physics. The change of title to *Theoretical and Mathematical Physics* (TMP) signals that the series is a suitable publication platform for both the mathematical and the theoretical physicist. The wider scope of the series is reflected by the composition of the editorial board, comprising both physicists and mathematicians.

The books, written in a didactic style and containing a certain amount of elementary background material, bridge the gap between advanced textbooks and research monographs. They can thus serve as a basis for advanced studies, not only for lectures and seminars at graduate level, but also for scientists entering a field of research.

Series Editors

Piotr Chrusciel, Wien, Austria

Jean-Pierre Eckmann, Genève, Switzerland

Harald Grosse, Wien, Austria

Antti Kupiainen, Helsinki, Finland

Hartmut Löwen, Düsseldorf, Germany

Kasia Rejzner, York, UK

Leon Takhtajan, Stony Brook, NY, USA

Jakob Yngvason, Wien, Austria

Peter Zoller, Innsbruck, Austria

Sergio Cecotti

Introduction to String Theory

Sergio Cecotti
Yanqi Lake Beijing Institute
of Mathematical Sciences and Applications
(BIMSA)
Beijing, China

ISSN 1864-5879 ISSN 1864-5887 (electronic)
Theoretical and Mathematical Physics
ISBN 978-3-031-36529-4 ISBN 978-3-031-36530-0 (eBook)
<https://doi.org/10.1007/978-3-031-36530-0>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2023

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To Magda and Luís

Preface

The textbook is a revised and enlarged version of the author's Lectures Notes on two-dimensional Conformal Field Theory and String Theory at SISSA (Trieste) and BIMSA (Beijing). The target audience is graduate students with basic mathematical background which does not encompass all the geometry which is required for a deep and elegant formulation of String Theory. Having this specific audience in mind, the author wrote a textbook with a “dual” structure: the treatment in the main text is kept elementary (with most technicalities omitted) but the text is accompanied by several “technical boxes” where the relevant mathematical definitions, theorems, and proofs are stated in precise terms, the concepts and constructions are clarified, and the theory is placed in its proper context, providing a broader (and deeper) perspective on the structure of String Theory. Some proofs are left to the reader as exercises. Shorter technical comments are given as footnotes. There are also special sections, marked ★, where additional material is presented; these sections may be skipped in a first reading. Technical/advanced stuff is confined in the appendices to each chapter.

The main focus of the textbook is *Superstring Theory*, that is, string models which are supersymmetric (in spacetime). Conformal Field Theory is described merely as a *tool* for String Theory, while the Bosonic String is studied for the didactical convenience of introducing the basic ideas and techniques in a much simpler set-up where we can focus on the fundamental issues without being taken astray by extra details. The bosonic and supersymmetric strings are then discussed together (rather than first the bosonic theory, and then the supersymmetric one) and the various aspects of the bosonic strings are described when we need them.

M-theory and F-theory are briefly mentioned in Chaps. 13 and 14.

Warning: *conciseness is not a goal in the book*. Materials are presented more than once, from different viewpoints, to convey the many-facet nature of String Theory.

Beijing, China

Sergio Cecotti

Introduction: *Why String Theory?*

Superstring Theory and its non-perturbative completions (M- and F-theory) enjoy a *unique status* in the twin realms of physical theories and mathematical constructs. It may be argued that it is *the* universal entity in both realms. For this reason, Superstring Theory is also known as the **Theory of Everything** (TOE).

Let us justify the above claim. Superstring Theory is a quantum system which is fully consistent and contains General Relativity in the sense that at low energies (equivalently at a large distance) it describes *inter alia* a massless spin-2 graviton which couples to the energy–momentum tensor of some low-energy effective matter system and obeys the Einstein equations of motion.

This fact is very remarkable on at least two counts.

First, it is well known that Quantum Physics and General Relativity are in severe tension; putting the two principles together leads to a number of conundrums and fundamental paradoxes that cannot be resolved unless several rather unlikely miracles happen. Consistency of gravity with quantum physics imposes on the theory a huge set of *sharp* consistency conditions which appears vastly overdetermined; the various physical quantities should be connected by *precise* relations which carry a distinct Number Theoretic flavor, in particular, a family of putative quantum-gravity theories which depends on a set of continuous parameters λ (“coupling constants” in the standard terminology) *cannot satisfy* the consistency conditions identically for all values of the λ ’s: they can be satisfied (at most) for some magic sharp value of the coupling parameters. In other words, if a consistent quantum theory of gravity exists at all, it is *fully rigid*.

Given the overdetermined nature of these conditions, one may ask whether the basic axioms of Quantum Physics and General Relativity are *inherently* incompatible so that no consistent quantum theory of gravity exists. A positive answer to this question would be a major blow to fundamental science; the real world does contain gravity and it looks quantum, so a positive answer will imply that there is a major mistake in our comprehension of either the basic principles of Quantum Physics or of the relation between gravity and the (large-scale) geometry of spacetime. Luckily, Superstring Theory provides us with several explicit examples of consistent models

of quantum gravity proving that the two physical principles are compatible albeit in a very subtle way.

This situation allows us to use (Super)String Theory (and its allied quantum systems) as a laboratory to study how the subtle consistency conditions work in detail, and to unravel the dynamical mechanisms beyond the unlikely miracles of Quantum Gravity. Thus, String Theory sheds light on the inner nature of Quantum Gravity in general, allowing us to infer the fundamental physical principles which underlie it.

“Rigidity” of Quantum Gravity, and the overdetermined nature of its consistency requirements, suggest that there are just a “few” distinct solutions to the conditions, that is, only a “few” consistent theories of Quantum Gravity each of which is rigid, i.e. without adjustable parameters. There is a widespread belief that only *finitely many* quantum gravities exist (albeit “phenomenologically” their number may look huge).

This entails that string theories provide a *finite fraction* of all quantum gravities. It may even be true that they exhaust the full class. Even if this is not the case, given the restrictiveness of the consistency conditions, one expects that there are “not many more” consistent quantum gravities and that all consistent models “look roughly like string theory” since they should realize the same “implausible” miracles. This idea is a (conjectural) physical principle called the “String Lamppost Principle” (SLP).

This state of affairs has dramatic consequences for physics in general. Let us say that a physical system is “realistic” if it can be materially realized in the real world, either in the laboratory or elsewhere in the universe. We say that a physical theory is *realistic* if it describes some “realistic” system in a physically realizable limit (such as low-energy or large-charge). Since the real world is quantum, any *realistic* physical theory can be completed into a quantum theory. Moreover, all material physical system has a non-zero weight, i.e. admits a coupling to gravity. We conclude that

all *realistic* physical theories describe (limits of) subsectors of some quantum gravity and so are “very similar” to subsectors of String Theory. In other words: the superstring is *essentially universal* in the category of realistic theories

By the last sentence we mean that a “sizeable fraction” of all realistic physical systems can be engineered in String Theory, while all realistic physical systems “look similar” to some systems engineered by String Theory. This follows from the SLP. Thus, in a sense, a topic in Theoretical Physics is either unrealistic (in the technical sense) or just a special subject in String Theory which is the Theory of Everything.¹

Another consequence of the overdetermined nature of the consistency conditions in Quantum Gravity is the existence of a large web of *dualities*. The point is that there are far more consistent math constructions and physical theories than consistent quantum systems. Hence several a priori unrelated models should describe *the same* fundamental physical object. The several constructions/theories describing the same quantum system are said to be related by a (quantum) duality. There are many different

¹ The two classes are not disjoint, there are “non-realistic” physical systems that can be engineered in string theory; these systems can be embedded in consistent theories of Quantum Gravity different from the one which describes our actual world (say with a lot of supersymmetries).

kinds of dualities in String Theory; the fundamental ones will be discussed in the book.

Several of the “usual” quantum systems, in particular, QFTs, have known constructions in String Theory. This is a consequence of the principle in the gray box. This also entails that the intrinsic dynamics of these QFTs are much more transparent when studied from the superstring perspective than with the usual field-theoretic methods. String Theory allowed a tremendous advance in our non-perturbative understanding of Quantum Field Theory. In particular, it produced interacting QFTs in more than four spacetime dimensions (such theories cannot be constructed by any conventional mean). See Chaps. 13 and 14 for basic examples.

The *swampland program*, initiated by Cumrun Vafa, aims to characterize the small class of physical systems which can be embedded in a consistent quantum theory of gravity out of the huge set of consistent-looking quantum systems. The goal is to get a list of universal necessary conditions that all “realistic” systems should obey. Here the lessons drawn from String Theory are crucial. We shall comment on the swampland conditions at various points in the book whenever we find a result in a concrete string model which uplifts to a (conjectural) universal property of Quantum Gravity.

The relation of String Theory to Quantum Gravity is remarkable in yet another way. Historically, String Theory was not constructed with the aim of formulating a theory of gravity. The original motivation for its introduction was to model hadronic spectroscopy. Ironically, when people discovered that String Theory describes a massless spin-2 particle, they thought it was a *failure* of the model (seen as a theory of the strong interactions). It was an unavoidable failure: that is, the existence of gravity is an *universal* and *automatic* prediction in String Theory which does not require any ad hoc construction or assumption. To date, String Theory is the only conceptual framework that makes this fundamental prediction without any ad hoc input. This fact suffices in itself to conclude that String Theory must be on the right track to lead us to the ultimate theory.

Contents

Part I Preliminary Matters

1	Introducing Strings: The Polyakov Path Integral	3
1.1	Introduction	3
1.1.1	History and <i>Cartoons</i>	3
1.1.2	Why String Theory?	5
1.1.3	String Theories: Geometric Classification	7
1.2	Bosonic String: The Polyakov Action	8
1.3	Bosonic String: Light-Cone Quantization	14
1.3.1	Quantization in the Light Cone	15
1.3.2	Lorentz Invariance: Emergence of Gravity	21
1.4	Covariant Quantization <i>à la</i> Polyakov	23
1.4.1	World-Sheet Topologies. Non-orientable Σ 's	24
1.4.2	Conformal Killing Vectors and Complex Automorphisms	24
1.4.3	The Fadeev–Popov Determinant	28
1.4.4	The Matter Sector	30
1.5	The Weyl Anomaly	31
1.5.1	★ Strings in Non-critical Dimensions	33
1.6	Ghost Zero-Modes: $\text{Aut}(\Sigma)$ and WP Moduli Geometry	33
1.6.1	The Riemann–Roch Theorem	36
1.6.2	b Zero-Modes and the Moduli Space	40
1.7	The Superstring	44
1.8	Strings Moving in Curved Backgrounds	50
1.8.1	The Spacetime Effective Action	55
1.8.2	String Compactifications	57
1.9	Physical Amplitudes, S -Matrix, and Vertices	58
	References	62

2	Review of 2d Conformal Field Theories	67
2.1	Spacetime Symmetries in QFT	67
2.2	Conformal Field Theory (CFT)	72
2.2.1	Conformal Automorphisms and Equivalences	73
2.2.2	Radial Quantization and the State-Operator Isomorphism	76
2.2.3	Operator Product Expansions (OPE)	79
2.3	CFT in 2d	81
2.3.1	Primary Fields	86
2.3.2	The Virasoro Algebra	89
2.3.3	Finite Conformal Transformation of $T(z)$	94
2.3.4	Representations on the Hilbert Space	95
2.3.5	Unitarity	99
2.3.6	General Chiral Algebras in 2d CFT	100
2.3.7	Partition Function and Modular Invariance	103
2.3.8	More on Correlation Functions. Normal Order	105
2.4	Example: The 2d Free Massless Scalar	108
2.5	Free SCFTs and Their Bosonization	113
2.5.1	b, c and β, γ Systems	114
2.5.2	Anomalous $U(1)$ Current (“Ghost Number”)	115
2.5.3	Fermi/Bose Sea States	116
2.5.4	The $U(1)$ Stress Tensor and Its Bosonization	119
2.5.5	Riemann–Roch and Bosonization: The Linear Dilaton CFT	122
2.5.6	Bosonization of β, γ : The $c = -2$ System	123
2.5.7	The Picture Charge	125
2.6	Inclusion of Boundaries: Non-orientable Surfaces	129
2.7	KaC–Moody and Current Algebras	133
2.7.1	KaC–Moody Algebras	133
2.7.2	The Sugawara Construction	137
2.7.3	★ Knizhnik–Zamolodchikov Equation	142
2.7.4	Simply Laced G at Level 1	145
2.7.5	Fermionic Realization of the Current Algebra	146
2.8	$(1, 1)$ Superconformal Algebra	147
2.8.1	Primary Superfields	149
2.8.2	Ramond and Neveu–Schwarz Sectors	151
2.8.3	SCFT State-Operator Correspondence	154
2.8.4	Example: The Free SCFT	158
2.9	$SO(2n)$ Current Algebra at Level 1 and Lattices	162
2.9.1	The $SO(d-1, 1)$ World-Sheet Current Algebra	162
2.9.2	Bosonization of the $SO(2N)$ Current Algebra	163
2.9.3	$\text{Spin}(8)$ Triality and Refermionization	170
2.10	On Classification of 2d Superconformal Algebras	172
2.10.1	★ Classification of 2d Superconformal Algebras	177
	Appendix 1: Witten’s Non-abelian 2d Bosonization	180

Appendix 2: Valued Graphs, Affine Lie Algebras, McKay	
Correspondence, and All That	185
References	190

Part II Constructing Superstring Theory

3 Spectrum, Vertices, and BRST Quantization	197
3.1 The Superstring Lorentz Current Algebra	197
3.2 The Physical Spectrum: Light-Cone Gauge	202
3.3 Old Covariant Quantization	206
3.4 OCQ: Physical Conditions Versus 2d Superfields	214
3.5 BRST Invariance: Generalities	218
3.6 BRST Quantization of the Bosonic String	222
3.7 BRST Quantization of the Superstring	227
3.7.1 Q -Homotopies: Picture Changing	231
3.7.2 BRST Cohomology in Operator Space: Vertices	238
3.7.3 RR Vertices and a Perturbative Theorem	241
3.8 Spacetime Supersymmetry	247
3.8.1 Supersymmetry Ward Identities: Absence of Tadpoles	249
3.9 Open Strings: Chan–Paton Degrees of Freedom	251
Appendix: Details on the No-Ghost Theorem	255
References	262
4 Bosonic String Amplitudes	265
4.1 Path Integrals for Non-compact Scalars	266
4.1.1 Scalar Amplitudes on World-Sheets with $\chi \geq 0$	269
4.2 Amplitudes for the b, c CFT	279
4.3 The Veneziano Amplitude	283
4.4 Chan–Paton Labels and Gauge Interactions	289
4.5 Closed String Tree-Level Amplitudes	291
4.5.1 Closed String Amplitudes on the Disk and $\mathbb{RP}^2$	294
4.6 One-Loop Amplitudes: The Torus	295
4.7 One-Loop: The Cylinder	301
4.8 Boundary and Cross-Cap States	304
4.9 One-Loop: Klein Bottle and Möbius Strip	305
References	308
5 10d Superstring Theories	309
5.1 2d Global Gravitational Anomalies	309
5.2 Consistent Closed Superstring Theories in 10d	320
5.3 Consistent Unoriented and Open Superstrings	324
5.4 2d Fermionic Path Integrals	328
5.5 Modular Invariance in Type II	335
5.6 Divergences and Tadpoles in Type I Theories	341
5.6.1 Consistency of $SO(32)$ Type I	346
References	350

6	Bosonic String: T-Duality & D-Branes	351
6.1	Toroidal Compactifications in Field Theory	351
6.2	2d CFT of a Compact Scalar	359
6.3	★ Bosonization: Riemann Identities for Partition Functions	367
6.4	T -Duality in Closed Strings	371
6.4.1	T -Duality for a Compact Scalar	375
6.4.2	★ T -Duality on a General Background and Buscher Rules	379
6.4.3	Compactification of Several Dimensions	381
6.5	Narain Compactifications	384
6.5.1	The T -Duality Group	389
6.6	Abelian Orbifolds	394
6.6.1	Twisting Procedure	399
6.6.2	★ More on the Kosterlitz–Thouless Transition Point	403
6.7	Open Strings: Adding Wilson Lines	404
6.8	Open Bosonic String: T -Duality	407
6.9	D -Branes	410
6.9.1	D -Brane Action (Bosonic String)	413
6.10	T -Duality of Unoriented Strings: Orientifolds	421
	Appendix 1: Non-Abelian Orbifolds	424
	Appendix 2: Classification of $c = 1$ CFTs	425
	References	426
7	The Heterotic String	429
7.1	Constructing String Models	429
7.2	The $SO(32)$ and $E_8 \times E_8$ Heterotic Strings in 10d	430
7.3	Non-supersymmetric Heterotic Strings in 10d	442
7.4	Heterotic Strings: The Bosonic Construction	448
7.5	Classification of Even Self-dual Lattices	453
7.6	SUSY Heterotic Strings in $d = 10$ (Bosonic Form)	458
7.7	Toroidal Compactifications	461
7.7.1	Relation Between $E_8 \times E_8$ and $SO(32)$ Heterotic Strings	469
7.7.2	Example: Toroidal Compactification to Four Dimensions	471
7.8	Supersymmetry and BPS States	473
	Appendix: The $N = 2$ Superstring	475
	References	478
Part III Physics of Supersymmetric Strings		
8	Low-Energy Effective Theories	483
8.1	Supergravity: a Quick Review	483
8.2	Non-Renormalization Theorems, BPS Objects	488

8.3	Supergravity in 11d	491
8.4	Type IIA Superstring: Low-Energy Effective Theory	494
8.5	Type IIB: Effective Low-Energy Theory	501
8.6	Type I Superstring: Low-Energy Effective Theory	506
8.7	Heterotic String	508
8.8	BPS Solutions	509
	References	522
9	Anomalies and All That	527
9.1	Review of the Anomaly Polynomial Formalism	527
9.2	Anomaly Cancellation in 10d SUSY String Theories	535
9.3	Modular-Invariant $\Rightarrow$ Anomaly-Free	542
	References	549
10	Superstring Amplitudes Non-Renormalization Theorems	551
10.1	Tree-Level Amplitudes	551
10.2	General Amplitudes	556
10.3	One-Loop Amplitudes	563
10.4	Non-Renormalization Theorems Again	572
	References	574
11	Calabi–Yau Compactifications	577
11.1	Geometric Background	577
11.1.1	Mini-Review of Differential Geometry (DG)	577
11.1.2	Complex and Kähler Manifolds	584
11.1.3	Calabi–Yau Manifolds (CY)	587
11.1.4	Ultra-short Review of Kodaira–Spencer (KS) Theory	593
11.2	Superstrings on CY Manifolds: The World-Sheet Perspective	597
11.2.1	Calabi–Yau 2d σ -Models as (2, 2) SCFTs	599
11.3	(2,2) SCFTs as Type II Backgrounds	613
11.4	Mirror Symmetry	617
11.5	Heterotic $E_8 \times E_8$ on a Calabi–Yau 3-fold	618
11.6	Type II Compactified on a 3-CY: the Spacetime Perspective ...	621
11.7	Lightning Review of 4d $\mathcal{N} = 2$ Supergravity	626
11.8	The Low-Energy Theory of Type IIB on a 3-CY X	634
11.9	The Hypermultiplet Sector. c -Map	644
11.10	Global Aspects	644
	References	645

Part IV Superstrings Beyond Weak Coupling

12	Superstring D-Branes	653
12.1	T -Duality in Type II Strings	653
12.2	T -Duality of Type I Strings: SUSY D-Branes	656
12.3	Relations Between Superstring Theories	661

12.4	D-Brane Tensions and RR Charges	663
12.5	D-Brane Actions	666
12.6	Supersymmetric Multi-brane Arrangements	672
12.6.1	Branes of Different Dimension Parallel to the Axes	672
12.6.2	The World-Volume Viewpoint: The $\#_{\text{ND}} = 4$ System	675
12.6.3	Non-parallel Branes	679
12.7	BPS Bound States of Branes	688
12.7.1	F1-D1 Bound States	689
12.7.2	D0-D p Bound States	700
12.8	D-Branes as Yang–Mills Instantons	710
	References	716
13	SUSY Strings at Strong Coupling	719
13.1	Type IIB Strings at Strong Coupling: $SL(2, \mathbb{Z})$ Duality	721
13.1.1	$SL(2, \mathbb{Z})$ Duality	724
13.1.2	D3-Branes and Montonen–Olive Duality	728
13.2	U -Duality	729
13.3	★ IIA on K3 is Dual to Heterotic on T^4	737
13.4	$SO(32)$ Type I-Heterotic Duality	739
13.4.1	The Type I D5-Brane Versus the Heterotic NS5	743
13.5	Type IIA at Strong Coupling: M-Theory	746
13.6	M-Theory BPS Objects Versus IIA Branes	750
13.7	The $E_8 \times E_8$ Heterotic String at Strong Coupling	761
13.8	IIA D8-Branes Versus M-Theory	764
13.9	The Big Picture: <i>What Is String Theory?</i>	765
	References	766
14	Applications and Further Topics	769
14.1	Taub-NUT and GH Geometries	769
14.1.1	Half-BPS 6-Branes in M-Theory. Non-Abelian Gauge Symmetry	776
14.2	ADHM Construction Versus D-Branes	779
14.3	The Idea of F-Theory	785
14.3.1	Duality Between M- and F-Theory	794
14.4	Matrix Theory: A Proposal for M-Theory	797
14.4.1	The M-Theory Membrane	800
14.5	6d (2, 0) SCFTs	806
14.5.1	Construction of 6d (2, 0) SCFT from IIB on $\mathbb{C}^2/\Gamma$	810
14.6	Quantum Physics of Black Holes	813
	References	818
	Index	823

Acronyms

(n) d	with (n) a positive integer number stands for (n) -dimensional; for instance, 2d stands for two-dimensional, 3d for three-dimensional, and so on
ADHM	Atiyah–Drinfeld–Hitchin–Manin (construction)
AdS	Anti-de Sitter (geometry, metric)
AG	Algebraic Geometry
a.k.a.	also known as
B	bosonic (string)
b.c.	boundary condition(s)
BH	Black Hole
BI	Born-Infeld-(Dirac) (action)
BPS	Bogomol’nyi–Prasad–Sommerfield (state, configuration, operator)
BRST	Becchi–Rouet–Stora–Tyutin (quantization)}
CCR	canonical commutation relation
cf.	compare, see
CFT	Conformal Field Theory
CKV	conformal Killing vector(s)
CKS	conformal Killing spinor(s)
CP	Chan–Paton (indices, labels, d.o.f.)
CS	Chern–Simons (term, coupling, invariant)
CW	Coleman–Weinberg (formula)
CY	Calabi–Yau (manifold, geometry, metric)
D	Dirichlet (boundary condition)
DDF	Del Giudice–Di Vecchia–Fubini (operators, states)
DG	Differential Geometry
d.o.f.	degree(s) of freedom
EDM	eventually distance minimizing (geodesics)
e.g.	for example, for instance, <i>exempli gratia</i>
e.o.m.	equations of motion
eq(s).	equation(s)
fig.	figure

FP	Fadeev–Popov (ghosts)
GH	Gibbons–Hawking (metrics, geometries)
GMN	Gaiotto–Moore–Neitzke
GR	General Relativity
GSO	Gliozzi–Serk–Olive (projection)
HJ	Hamilton–Jacobi (formulation, action)
HK	hyperKähler (geometry, metric, manifold)
i.e.	that is, <i>id est</i>
IPR	infinitesimal period relations
KK	Kaluza–Klein (metric, mechanism, modes)
KM	Kač–Moody (algebra)
KS	Kodaira–Spencer (theory, vector)
IR	Infra-red, alias the extreme low-energy limit
l.c.	light cone (gauge, Hilbert space, spectrum)
LHS	left-hand side (of an equation)
MCQ	Mapping Class Group
MUM	Maximally unipotent monodromy
MW	Majorana–Weyl (spinor)
N	Neumann (boundary condition)
NG	Nambu–Goto (action)
NS	Neveu–Schwarz (sector of the superstring)
NS-R	Neveu–Schwarz/Ramond (formulation of the superstring)
NT	Number Theory
OCQ	Old covariant quantization
ODE	Ordinary differential equation(s)
OPE	Operator product expansion
PCO	Picture changing operator
PDE	Partial differential equation(s)
QCD	Quantum chromodynamics
QED	Quantum electrodynamics
QFT	Quantum Field Theory
QK	quaternionic Kähler (geometry, metric, manifold)
QM	Quantum Mechanics (i.e. QFT in 1d)
R	Ramond (sector of the superstring)
RCFT	Rational conformal field theory
RG	Renormalization group
rhs	right-hand side (of an equation)
repr(s).	representation(s)
resp.	respectively
SBI	super-Born-Infeld (theory, Lagrangian)
SCFT	Superconformal Field Theory
SLP	String lamp-post principle
SQM	Supersymmetric Quantum Mechanics
SRS	Super Riemann surface

SUGRA	Supergravity, i.e. a supersymmetric theory containing General Relativity
SUSY	Supersymmetry (noun) or supersymmetric (adjective)
SYM	Super–Yang–Mills
TFT	Topological Field Theory
TOE	Theory of Everything
tr	transverse (adjective)
UV	Ultra-violet, alias the extreme high-energy limit
v.e.v.	vacuum expectation value
VHS	variation(s) of Hodge structure
vs.	versus
WP	Weil–Petersson (geometry, metric, etc.)
WZW	Wess–Zumino–Witten (model, action)
YM	Yang–Mills (theory, symmetry, coupling)

Conventions and Symbols

General Conventions

- If the symbol a stands for an object which is left-moving on the (oriented) string world-sheet, the symbol $\tilde{a}$ with a tilde over it stands for the corresponding right-moving object. In other words, world-sheet parity acts on all symbols as the involution $a \leftrightarrow \tilde{a}$.
- **Warning.** The above convention has two *exceptions*:
 - a tilde $\tilde{M}$ over the symbol M for a geometric space stands for the universal cover of the space M
 - a tilde $\tilde{F}^{(k+1)}$ over the symbol $F^{(k+1)}$ of the field strength of a k -form gauge field $C^{(k+1)}$ means that the field strength is improved
- if $G, H, I, J, \dots, SU(N), SO(N), Sp(N), G_2, \dots$ are Lie groups, their Lie algebras are denoted by the corresponding lower-case German letters: $\mathfrak{g}, \mathfrak{h}, \mathfrak{i}, \mathfrak{j}, \dots, \mathfrak{su}(N), \mathfrak{so}(N), \mathfrak{sp}(N), \mathfrak{g}_2, \dots$

Symbols

Linear Algebra

$\mathbb{F}(k)$	($\mathbb{F}$ a field or a ring) the algebra of $k \times k$ matrices with entries in $\mathbb{F}$
$\mathbf{1}_n$	the $n \times n$ identity matrix
$\oplus$	direct sum (of vector spaces, modules, representations, etc.)
$A^{\oplus n}$	direct sum of n copies of the object A
$\otimes$	tensor product (of vector spaces, modules, representations, etc.)
$\otimes^k$	tensor k -th power
$\wedge$	antisymmetric tensor product, external product

$\otimes^k$	antisymmetric tensor k -th power
$\odot$	symmetric tensor product (of vector spaces, representations, etc.)
$\odot^k$	symmetric tensor k -th power
$V^\vee$	the dual of V over the ground field (V a vector space, a representation, etc.)
γ_μ, Γ_m	(“Dirac”) matrices representing the generators of a Clifford algebra

Groups & Lie Algebras

$ G $	order of the group G , i.e. the cardinality of the underlying set
$\langle \xi_1, \dots, \xi_n \rangle$	group generated by elements $\xi_1, \dots, \xi_n$
$\langle \xi_1, \dots : R_1, \dots \rangle$	group generated by $\xi_1, \dots, \xi_n$ subjected to the relations $R_1, \dots, R_m$
$\mathbb{Z}_n$	the cyclic group with n elements $\mathbb{Z}/n\mathbb{Z}$
$\mathfrak{S}_n$	symmetric group in n letters
$\mathcal{B}_n$	braid group in n strands
$\text{Weyl}(\mathfrak{g})$	Weyl group of the semi-simple Lie algebra $\mathfrak{g}$
$\Delta(\mathfrak{g})$	set of roots of the Lie algebra $\mathfrak{g}$
$r_{\mathfrak{g}}$	rank of the Lie algebra $\mathfrak{g}$
α_i	the simple root associated to the i -th node of the Dynkin graph
$SL(n, \mathbb{F})$	the (Lie, algebraic, or arithmetic) multiplicative group of matrices in $\mathbb{F}(n)$ with determinant 1
$SO(p, q, \mathbb{F})$	the subgroup of $SL(n, \mathbb{F})$ preserving a symmetric quadratic form of signature (p, q)
$Sp(2k, \mathbb{F})$	the subgroup of $SL(2k, \mathbb{F})$ preserving a non-degenerate skew-symmetric quadratic form
Weyl	group of Weyl rescaling of the metric in a Riemannian manifold
Diff (M)	diffeomorphism group of the smooth manifold M . M may be omitted when there is no danger of confusion.
Diff $^+(M)$	subgroup of Diff $^+(M)$ (for M oriented) preserving orientation
Diff $^0(M)$	normal subgroup of Diff $^+(M)$ (for M oriented and connected) of elements homotopic to the identity
MCG (M)	mapping class group of the manifold M . For M oriented and connected, $\text{MCG}(M) \equiv \text{Diff}^+(M)/\text{Diff}^0(M)$

Particular spaces

$\mathbb{F}^n$	the vector (affine) space over $\mathbb{F} = \mathbb{R}, \mathbb{C}$ of dimension n
$\mathbb{R}^{r,s}$	$\mathbb{R}^{r+s}$ endowed with a flat pseudo-Riemannian metric $\eta_{\mu\nu}$ of signature (r, s)
S^k	the k -dimensional round sphere $S^k \equiv \{x^i \in \mathbb{R}^{n+1} : \sum_i (x_i)^2 = 1\}$
T^k	the k -dimensional torus $(S^1)^k$

$\mathbb{P}^n$	the complex projective space of dimension n . In particular: $\mathbb{P}^1 \simeq S^2$ the <i>Riemann sphere</i>
$\mathbb{RP}^n$	the real projective n -space, that is, the space of real-valued points in the projective n -space $\mathbb{P}^n$
Σ	the string world-sheet: a manifold of real dimension 2 (complex dimension 1)
$\check{\Sigma}$	the oriented double of the non-orientable 2-manifold Σ
Σ_g	an oriented 2-manifold (or a Riemann surface) of genus g
$\mathbb{C}^\times$	the punctured complex plane $\mathbb{C}^\times \equiv \{z \in \mathbb{C} : z \neq 0\}$ conformal to the infinite cylinder $\mathbb{R} \times S^1$
$\mathcal{H}$	the upper-half plane $\mathcal{H} \equiv \{z \in \mathbb{C} : \text{Im} z > 0\} \subset \mathbb{C}$ in the complex plane $\mathbb{C}$
$\mathcal{D}$	the open unit disk $\mathcal{D} \equiv \{z \in \mathbb{C} : z < 1\} \subset \mathbb{C}$ in the complex plane $\mathbb{C}$
Δ^*	the punctured unit disk $\Delta^* \equiv \{z \in \mathcal{D} : z \neq 0\}$
St	the infinite strip $\text{Cy} \equiv [0, \pi] \times \mathbb{R}$
Cy	the finite cylinder $\text{Cy} \equiv [0, \pi] \times S^1$. Also called the <i>annulus</i>
Mö	the Möbius strip
Kl	the Klein bottle

Manifolds & Calculus

∂M	boundary of the manifold M
TM	smooth tangent bundle of a smooth manifold M or the holomorphic tangent bundle of a complex manifold M
T^*M	cotangent bundle dual to TM
$\Omega^k(M)$	bundle of smooth k -forms on the manifold M , $\Omega^k(M) \equiv \wedge^k T^*M$
$\Omega^\bullet(M)$	exterior algebra $\oplus_k \Omega^k(M)$
$\Lambda^{p,q}(M)$	bundle (or sheaf) of smooth (or holomorphic) forms of type (p, q) on the complex manifold M
$\mathcal{O}$	sheaf of germs of holomorphic functions a.k.a. the structure sheaf
$\mathcal{O}^\times$	sheaf of germs of nowhere vanishing holomorphic functions
$\mathcal{L}$	holomorphic line bundle (usually on a Riemann surface)
$*$	the Hodge dual in a (pseudo)Riemannian manifold
d	exterior derivative
δ	formal Hermitian dual of d . $\delta \equiv -*/, d*$
ι_v	contraction (of a differential form) with the vector field v
$\mathcal{L}_v$	Lie derivative along the vector field v ($\mathcal{L}_v = d\iota_v + \iota_v d$ acting $\Omega^\bullet(M)$)
D_a	covariant derivative (with respect to some connection)
∇_a	covariant derivative (with respect to the Levi-Civita connection)
$\not{D}$	covariant Dirac operator $\gamma^\mu D_\mu$
$\square$	Laplacian ($\square = d\delta + \delta d$ acting $\Omega^\bullet(M)$)

Riemann Surfaces

g	genus of the surface (number of handles)
b	number of connected components of the boundary
χ	Euler characteristic
K	canonical bundle
$\mathcal{M}_g$	moduli space of complex structures of a genus g surface
$\mathcal{M}_{g,n}$	complex moduli space of a genus g surface with n punctures

Fields

Φ	a short-hand symbol for all the fields in the theory
X^μ	2d scalars, embedding coordinates of the string
ψ^μ	2d left-moving fermions supersymmetry partner of X^μ
b, c	2d Fadeev–Popov ghosts of reparametrization invariance
β, γ	2d Fadeev–Popov ghosts of local supersymmetry
ϕ	scalar field which bosonize the β, γ current
λ^A	2d left-moving fermions (heterotic string)
$\phi^a(z)$	2d chiral scalars
$g_{\mu\nu}, G_{\mu\nu}$	target space metric
$B_{\mu\nu}$	target space NS-NS 2-form field
$H_{\mu\nu\rho}$	3-form field strength of $B_{\mu\nu}$
Φ	target space dilaton
A_μ	target space (1-form) gauge field
$F_{\mu\nu}$	2-form field strength of A_μ
$C^{(k)}$	k -form gauge field
$F^{(k+1)}$	$(k + 1)$ -form field strength of $C^{(k)}$

2d Conformal Field Theory

$\mathfrak{A}$	algebra of local operators
$T(z), T_B(z)$	(left-moving) energy–momentum tensor
L_m	modes of the energy–momentum tensor
c	Virasoro central charge
h	Virasoro weight
$T_F(z)$	(left-moving) supercurrent
G_r	modes of the supercurrent
$\hat{c}$	$2c/3$
α_n	modes of the free scalar field
$J(z)$	left-moving chiral current
$S_\alpha(z)$	spin fields

List of Off-Text Technical Boxes

BOX 1.1	Properties of 2d manifolds of Lorentzian signature	15
BOX 1.2	Casimir energies for free massless fields on S^1	20
BOX 1.3	Lorentz symmetry, critical dimension, and K3 surfaces	23
BOX 1.4	Local and global structure of 2d Riemannian manifolds	25
BOX 1.5	Topological classification of closed 2-manifolds	26
BOX 1.6	Basic facts about line bundles on Riemann surfaces	37
BOX 1.7	Hermitian line bundles as $U(1)$ gauge fields	39
BOX 1.8	Rudiments of Kodaira–Spencer theory [63]	42
BOX 1.9	Uniformization of Riemann surfaces [59]	43
BOX 1.10	Spin structures on Riemann surfaces [79]	47
BOX 1.11	Superconformal symmetries of the 2-sphere	49
BOX 1.12	Conformal properties of $R_{\mu\nu}$ and R in d dimensions	56
BOX 1.13	Extra dimensions: are they “real”? (Heuristics)	58
BOX 2.1	More on the Euclidean conformal group $SO(d + 1, 1)$	74
BOX 2.2	Holomorphic line bundles versus divisors	85
BOX 2.3	Weyl anomaly versus Virasoro central charge c	92
BOX 2.4	Complex moduli of tori and modular invariance	102
BOX 2.5	No spontaneous symmetry breaking in 2d	109
BOX 2.6	Ghost current anomaly versus Riemann–Roch theorem	122
BOX 2.7	An alternative bosonization of the β, γ system	125
BOX 2.8	Coxeter number and dual Coxeter number	139
BOX 2.9	OPEs for the Sugawara energy–momentum tensor	140
BOX 2.10	Lattices: basic definitions	165
BOX 2.11	The Lie superalgebra $D(2 1; \alpha)$	181
BOX 3.1	The representation ring of $\text{Spin}(2n)$	204
BOX 3.2	Proof of Lemma 3.2	226
BOX 3.3	The Kähler-Dirac equation & BRST-invariant RR vertices	246
BOX 4.1	The scalar Green’s functions on the torus	273
BOX 4.2	Properties of the 2-2 S -matrix	286

BOX 4.3	Veneziano amplitude versus unitarity and critical dimension	287
BOX 4.4	Computation of the Virasoro–Shapiro integral	293
BOX 4.5	Poincaré geometry of the upper half-plane	299
BOX 4.6	Point-particle vacuum amplitude	301
BOX 5.1	Artin groups, Weyl groups, and Dynkin graphs	314
BOX 5.2	Mutual locality of world-sheet operators	320
BOX 5.3	<i>Spin & Statistics</i> theorem in physical 10d spacetime	321
BOX 5.4	Spacetime supercharges in the open string sector	326
BOX 5.5	ϑ -functions with characteristics	331
BOX 5.6	Modular properties of Z_{β}^{α}	332
BOX 5.7	Ω -twisted fermionic partition functions	334
BOX 5.8	An alternative viewpoint on the RR tadpole	345
BOX 5.9	Ω -twisted bosonic open string partition functions	348
BOX 6.1	Non-Abelian Kaluza–Klein geometries	355
BOX 6.2	Abelian gauge transformation of $B_{\mu\nu}$, Eq. (6.33)	358
BOX 6.3	Conservation of winding number	362
BOX 6.4	Poisson summation formula	363
BOX 6.5	The $SL(2, \mathbb{Z})$ Hodge norm	364
BOX 6.6	Winding number as the 2-form gauge charge	373
BOX 6.7	Riemannian symmetric spaces of non-compact type	390
BOX 6.8	Geometry of the symmetric space $O(k, k) / [O(k) \times O(k)]$	391
BOX 6.9	T -duality revisited	393
BOX 6.10	Volume of arithmetic quotients & Narain moduli spaces	395
BOX 6.11	Some useful identities between ϑ -functions	400
BOX 6.12	The scalar Green’s function in d -dimensions	419
BOX 7.1	All $(h, \bar{h}) = (1, 0)$ Primaries Generate Global Symmetries	441
BOX 7.2	Bernoulli Numbers, ζ -Function, and All That	455
BOX 7.3	The Order of the Weyl Group	457
BOX 7.4	Proof of the Duality (7.169)	469
BOX 8.1	The improved field strength $\tilde{F}^{(4)}$	496
BOX 8.2	Proofs of Eqs. (8.36) and (8.37)	497
BOX 8.3	‘Basic’ IIB SUGRA: further details	502
BOX 8.4	Eq. (8.99)	507
BOX 8.5	Eqs. (8.107) and (8.108)	508
BOX 8.6	Curvatures in warped products	511
BOX 8.7	1/2-BPS 2-Branes in 11d SUGRA	516
BOX 9.1	The anomaly polynomial is the index density in $d + 2$	532
BOX 9.2	Indices for the Kähler–Dirac operator $d + \delta$	535
BOX 9.3	Modular properties of ϑ -functions with characteristics	545
BOX 9.4	The modular function $j(\tau)$ as a partition function	548
BOX 10.1	Spurious poles in $\delta(\beta)$, $\delta(\gamma)$, $H(\beta)$ correlators	561
BOX 10.2	The generating function of integrated current correlators	566

BOX 10.3	Spin-structures and pictures in Eq. (10.85)	570
BOX 11.1	Families of Complex Manifolds and Monodromy	591
BOX 11.2	Frobenius Algebras	603
BOX 12.1	More on the product in Clifford algebras	655
BOX 12.2	Coulomb branch of maximal super-Yang-Mills	668
BOX 12.3	Expanding the action (12.45)	670
BOX 12.4	Some useful formulae for θ -functions	684
BOX 12.5	Proof of (12.118)	686
BOX 12.6	Equation (12.146)	691
BOX 12.7	The BPS bound is saturated	692
BOX 12.8	't Hooft twisted boundary conditions and all that	696
BOX 12.9	Properties of β matrices	701
BOX 12.10	More on the cohomology of the Hilbert schemes $X^{[n]}$	708
BOX 12.11	Sign of long-distance force between a D0 and a D2k	709
BOX 12.12	HyperKähler quotient	715
BOX 13.1	Couplings in Type IIB	723
BOX 13.2	Proof of Eq. (13.87)	753
BOX 13.3	Facts about 6d supersymmetric QFT	756
BOX 14.1	BPS monopole equations <i>versus</i> hyperholomorphic bundles	773
BOX 14.2	Flat metric in $\mathbb{R}^4$ as a Gibbons-Hawking metric	774
BOX 14.3	Properties of Euclidean 4d Weyl σ -matrices	780
BOX 14.4	$SU(N)$ instantons over $\mathbb{P}^1(\mathbb{H})$	781
BOX 14.5	Homological invariant and Kodaira monodromies	789
BOX 14.6	Kodaira fibers	793
BOX 14.7	Super-membrane in light-cone gauge	801
BOX 14.8	Quantum torus algebra and all that	804

Part I

Preliminary Matters

The Part I consists of two chapters.

Chapter 1 presents a general survey of perturbative (super)string theory from the viewpoint of the Polyakov integral over surfaces with or without fermions. The last two sections give a general discussion of the relation of Polyakov amplitudes to observables in physical spacetime.

Chapter 2 contains an introduction to 2d CFT. The theory is developed having in mind its applications to string theory: the treatment of topics crucial for (super)string theory is rather detailed; the other important aspects are either abridged or confined in advanced ★-sections and appendices.

Chapter 1

Introducing Strings: The Polyakov Path Integral



Abstract The first chapter serves dual purposes: *first* to give a general overview of perturbative string theory skipping all details and technicalities; *second* to describe the covariant quantization of bosonic and fermionic strings in terms of the Polyakov path integral over surfaces. The relevant geometrical facts are reviewed, while the Faddeev–Popov ghosts and their zero-modes are studied in great detail. The efficient tool to compute physical amplitudes in these theories is two-dimensional conformal field theory which is the subject of the following chapter. In the last two sections, we discuss the relation of the quantum amplitudes computed by the Polyakov integral with physical observables in possibly curved spacetimes and non-trivial backgrounds.

General references for string theory include [1–10].

1.1 Introduction

Contrary to mathematics, physics is usually taught “historically” in order to develop “physical intuition”. Unfortunately, the history of string theory is rather peculiar and of little didactical use. For detailed accounts of its early stages, see [11–13].

1.1.1 History and Cartoons

String theory was introduced in “pre-historical” times¹ as a phenomenological description of strong interactions. The rough picture was that hadrons are made of quarks carrying flavor and color degrees of freedom (d.o.f.), the flavor d.o.f. being “visible” and the color ones “confined”. A meson was thought of as a quark–anti-quark pair connected by a color flux-tube: such a configuration may be schematically

¹ The sixties of the twentieth century.

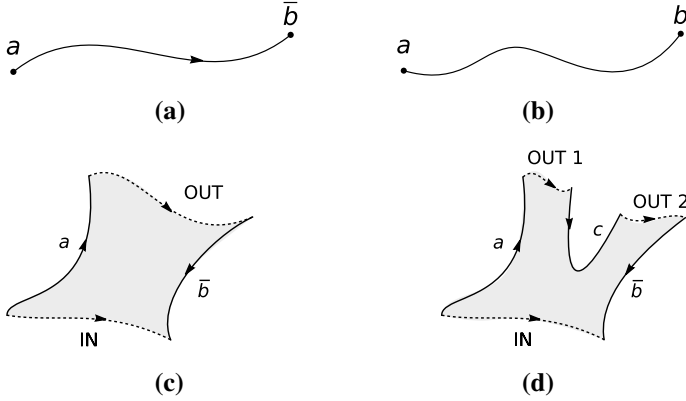


Fig. 1.1 **a** An *oriented* open string with labels $a, \bar{b}$ at its endpoints. **b** An *unoriented* open string with labels a, b at its endpoints. **c** The time evolution from the open string configuration **IN** to the string configuration **OUT** spans a two-dimensional surface in spacetime: the *world-sheet*. The boundaries of the world-sheet carry the labels $a, \bar{b}$. **d** A one-open-string state with endpoint labels $(a, \bar{b})$ evolving into a state containing two open strings with endpoint labels $(a, \bar{c})$ and $(c, \bar{b})$. The world-sheet has a new boundary component with label c

visualized as a curve embedded in spacetime with two ends labeled by indices corresponding to the flavor d.o.f. of the quarks;² see Fig. 1.1a.

The curve (called a *string* or *flux-tube*) was necessarily a dynamical object with its own degrees of freedom describing its motions and changes of shape with time.

Geometrically, there are two possibilities: either the curve is *oriented* (Fig. 1.1a) or *non-oriented* (Fig. 1.1b). In the second case, the two possible orientations of the curve describe the *same* physical state. Let G be a symmetry group which acts on the configurations of a quantum system: *gauging* the symmetry G amounts to declaring that all configurations in each G -orbit are *the same* physical state (no observable can distinguish them). Thus, non-oriented strings are obtained from the oriented ones by gauging the group $\mathbb{Z}_2$ of orientation flips. We write Ω for the generator of $\mathbb{Z}_2$, i.e. the inversion of orientation. Ω acts on the flavor indices by interchanging the labels at the two ends: to fully determine the system, we have to specify whether the flavor representation of the string is *symmetric* or *antisymmetric* under this operation Ω .

The world-story ($\equiv$ time evolution) of such a string describes a two-dimensional world-sheet Σ immersed in spacetime whose boundaries $\partial\Sigma$ are the world-lines of the “quarks” at its two ends; see Fig. 1.1c. The surface Σ is oriented (resp. non-oriented) if the string is oriented (resp. non-oriented). Interactions between mesons correspond to world-sheets with more boundaries: Fig. 1.1d represents a meson with flavor indices $(a\bar{b})$ decaying in two mesons of flavor $(a\bar{c})$ and $(c\bar{b})$.

In addition to mesons, we may consider closed flux-tubes with the topology of the circle S^1 . Their world-stories are given by world-sheets without boundaries: for

² Technically, each end of the string carries an index a (resp. $\bar{a}$) labeling the basis elements of the quark (resp. conjugate quark) representation space of the flavor Lie group.

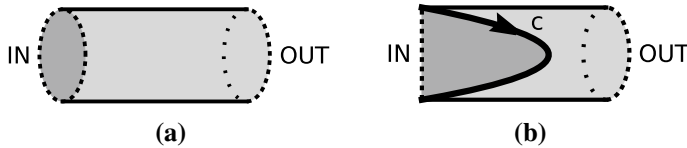


Fig. 1.2 **a** The world-sheet describing the world-story of a closed string propagating in spacetime has the topology of a cylinder. **b** The world-story of an open string evolving into a closed string. The boundary of the world-sheet is drawn thicker: it carries the label c

instance, the “free” propagation of a closed flux-tube sweeps an infinite cylinder (see Fig. 1.2a). A simple picture shows that we can produce closed flux-tubes out of mesons (i.e. out of flux-tubes ending in quarks), see Fig. 1.2b, this process is the time-reversal of a closed string breaking open. Hence, *the theory of open strings (“mesons”) does not exist by itself*: unitarity requires the existence of another sector consisting of closed flux-tubes (“glueballs”). Consistency requires the closed flux-tubes to be oriented (resp. non-oriented) if the mesonic flux-tubes (*open strings*) are oriented (resp. non-oriented). On the contrary, closed flux-tubes (either oriented or non-oriented) may exist by themselves, without other sectors.

This phenomenological model had some rough validity. For instance, one may show that the *exact* solution to $SU(N)$ QCD in two-dimensions, in the limit of large N , is given precisely by the above picture [14]. The 4d QCD flux-tube is instead described by a rather complicated non-fundamental string theory whose world-sheet action contains several ad hoc non-universal interactions. Thus, the stringy approach led to a description of strong interactions which is both intricate and of limited value, since it applies only up to a certain fundamental scale Λ_{QCD} . People eventually switched to QCD (and the Standard Model) as a more fundamental, more elegant, and much “simpler” description of strong interactions.

String theory disappeared from the radars for more than a decade. Eventually, it came back as *the* fundamental physical theory, also called the **Theory of Everything** (TOE). Let us explain the logic beyond this spectacular change of attitude.

1.1.2 Why String Theory?

In this textbook, we are interested in *fundamental* quantum strings, not phenomenological ones. Toward the end of the “pre-historical” era, it was realized that any *fundamental* theory of strings describes *inter alia* a massless spin-2 particle. In an (interacting) local relativistic theory, the presence of a massless spin-2 particle is consistent³ only if the spin-2 field is identified with the spacetime metric $g_{\mu\nu}$ and the theory itself is invariant under arbitrary reparametrizations of spacetime. The

³ This statement follows, for example, from the Weinberg–Witten theorem [15] or more technically from Weinberg’s S -matrix theorems [16–18].

low-energy physics is then effectively described by Einstein’s gravity coupled to some matter system. In other words: a fundamental string theory necessarily contains Einstein gravity. Since string theory is a quantum system from the very start, we conclude the following.

A fundamental quantum, relativistic, string theory is in particular a fully consistent theory of quantum gravity.

To date, strings are the only (known) solid way to construct a quantum consistent theory of gravity. Other approaches to quantum gravity exist, but their quantum consistency is not really established. Since the real world is quantum and contains gravity, at present string theory is the *only* hopefully realistic theory at our disposal. String theory may be not the “right” theory describing the universe we live in, but every other theory we are able to formulate at present—*especially* quantum field theory (QFT)—is most certainly incomplete or wrong.

Let me make this point sharper. The usual attitude (a.k.a. particle *phenomenology*) is that gravity is largely irrelevant in the description of most physical phenomena—such as collider physics—as long as they involve energies much below the Planck scale ($\approx 10^{19}$ GeV). This way of thinking has led to the prejudice that for “practical purposes” we can limit ourselves to “non-gravitational” quantum systems such as the so-called Standard Model. However this conclusion is not correct: any *real* physical system has a non-zero weight,⁴ i.e. must couple to gravity. This gives a strong selection rule on the class of physical theories which may claim to be realistic:

Any realistic quantum system should admit a quantum consistent coupling to quantum gravity; in particular, any realistic low-energy quantum field theory should admit a UV completion which is a quantum gravity.

This is a formidable constraint on “phenomenology”: indeed *almost all* “realistic looking” effective field theories do not have such a UV completion. This fact is easy to establish: the “realistic looking” consistent QFT are uncountably many (since they contain continuous free parameters), while theories which may be coupled to quantum gravity are at most countably many.⁵ Therefore the only known method

⁴ More dramatically (and technically): all quantum systems/states which can be realized in the laboratory can also be thrown inside a Black Hole. This process should avoid all information paradox—this cannot happen unless some very “fine-tuned” miracle takes place in the quantum system which falls in the Hole.

⁵ It is widely believed that the Standard Model admits a quantum gravity completion only for a few *sharp* values of the various couplings (such as the Yukawa ones). For instance: it is conceivable that the “inexplicable” phenomenological fact that leptons appear in 3 families, which are identical in every aspect except for their masses $m_\mu/m_e = O(10^2)$, $m_\tau/m_\mu = O(10^2)$, follows from a consistency requirement of the coupling to gravity.

to *prove* that a phenomenological model is realistic is to devise it as a subsector of string theory.

While (at the moment) we cannot say for certain that *all* realistic quantum systems are subsectors of string theory, there is a growing evidence that this is indeed the case. String theory is likely to be the *only* theory in physics—all other “theories” being just *special topics* in this unique theory.

1.1.3 String Theories: Geometric Classification

Although the fundamental strings behave quite differently from the naive flux-tube description of hadrons, the cartoonish description of their world-sheets remains valid since it depends only on basic facts of two-dimensional topology. We conclude that we can distinguish *four* geometric classes of string theories:

closed oriented the world-sheets are oriented⁶ 2-manifolds⁷ without boundaries;
closed unoriented the world-sheets are non-oriented 2-manifolds without boundaries;

closed plus open oriented the world-sheets are oriented 2-manifolds, possibly with boundaries, while each connected component of the boundary carries a discrete degree of freedom (an index taking N values) called the *Chan–Paton* (CP) index (or label);⁸

closed plus open unoriented the world-sheets are non-oriented 2-manifolds, possibly with boundaries, while the boundaries carry a discrete CP degree of freedom. The states of the open string are either symmetric or antisymmetric in the Chan–Paton indices.

Note 1.1 The boundary $\partial\Sigma$ of the world-sheet Σ is a disjoint union of b copies of the circle S^1 called the (*connected*) *components* of the boundary. If Σ is oriented, the 1-manifold $\partial\Sigma$ inherits an induced orientation.

The physical classification of string theories depends in addition on the actual d.o.f. which propagate along the string world-sheet. We start from the very simplest possibility: the *bosonic string*.

⁶ By definition an oriented world-sheet Σ is a *pair* $(\hat{\Sigma}, \varepsilon)$ where $\hat{\Sigma}$ is an orientable 2-manifold, while ε is a choice of orientation. We stress that for an oriented string the world-sheets $(\hat{\Sigma}, \varepsilon)$ and $(\hat{\Sigma}, -\varepsilon)$ describe physically inequivalent world-stories.

⁷ Here and throughout the book, “ k -manifold” stands for “smooth manifold of real dimension k ”.

⁸ In the “pre-historical” picture of hadrons the CP label is the flavor $U(N)$ group index.

1.2 Bosonic String: The Polyakov Action

From Nambu–Goto to Polyakov

We consider a string moving in flat d -dimensional Minkowski space M with coordinates X^μ and metric $\eta_{\mu\nu}$ ($\mu, \nu = 0, 1, \dots, d-1$).⁹ The spacetime M where the string moves will also be called the *target space*.

The time evolution ($\equiv$ world-story) of a string moving in M is described by the immersion map of its world-sheet Σ in spacetime

$$X: \Sigma \rightarrow M, \quad (\sigma^1, \sigma^2) \mapsto X^\mu(\sigma^1, \sigma^2), \quad (1.1)$$

where σ^1, σ^2 are local coordinates on the world-sheet Σ . The world-story depends only on image $X(\Sigma) \subset M$ of the map X and not on the way we parametrize it.

We look for an intrinsic, i.e. 2d reparametrization-invariant, description of the dynamics.¹⁰ A natural 2d parametrization-invariant action is the *Nambu–Goto* (NG) one, $S_{\text{NG}}[X]$, proportional to the area $\mathbf{A}[X(\Sigma)]$ of the world-sheet as an immersed submanifold $X(\Sigma) \subset M$ of spacetime, equipped with the induced 2d metric

$$h_{ab} \equiv \eta_{\mu\nu} \partial_a X^\mu \partial_b X^\nu \quad (1.2)$$

(here $\partial_a X^\mu \equiv \partial X^\mu / \partial \sigma^a$ and $a, b = 0, 1$), that is,

$$S_{\text{NG}}[X] = T \cdot \mathbf{A}[X(\Sigma)] \equiv T \int_{\Sigma} d^2\sigma \sqrt{-\det[h_{ab}]}. \quad (1.3)$$

The overall constant T is the *string tension*, traditionally written as

$$T = \frac{1}{2\pi\alpha'} \quad (1.4)$$

with α' called the *Regge slope*. The tension T has dimension $[\text{mass}]^2$. $\ell \equiv \sqrt{\alpha'}$ is a fundamental length scale called the *string length*.

The Nambu–Goto action is *classically* equivalent to the action

$$S[g, X] = \frac{1}{4\pi\alpha'} \int_{\Sigma} d^2\sigma \sqrt{-g} g^{ab} \eta_{\mu\nu} \partial_a X^\mu \partial_b X^\nu, \quad (1.5)$$

⁹ We use the “mostly +” convention for the Minkowski metric $(-, +, +, \dots, +)$.

¹⁰ Parametrization invariance is actually an *assumption*: it corresponds to the physical intuition of what a *fundamental* string should be. This is the straightforward generalization to the string (a one-dimensional object) of the standard description of the *fundamental* point particle (whose world-story is a curve $C \subset M$) which leads (via second quantization) to ordinary QFT. In the point-particle case, the action is proportional to the length of the world-line C . The area of the world-sheet is the most natural generalization of the length of the world-line.

where now g_{ab} is a “dynamical” 2d (Minkowskian) metric on the world-sheet Σ . Indeed the equations of motion of the 2d metric are

$$0 = \frac{2\pi\alpha'}{\sqrt{-g}} \frac{\delta S[g, X]}{\delta g^{ab}} = \partial_a X^\mu \partial_b X_\mu - \frac{1}{2} g_{ab} g^{cd} \partial_c X^\mu \partial_d X_\mu \equiv h_{ab} - g_{ab} (g^{cd} \partial_c X^\mu \partial_d X_\mu), \quad (1.6)$$

so that on-shell g_{ab} and h_{ab} differ only by an (arbitrary) overall factor

$$g(\sigma)_{ab} = e^{\phi(\sigma)} h(\sigma)_{ab}, \quad (1.7)$$

that is, by a local *Weyl rescaling*. The Weyl factor $e^{\phi(\sigma)}$ cancels in the combination $\sqrt{-g} g^{ab}$ which appears in the action (1.5) so that

$$\sqrt{-g} g^{ab} = \sqrt{-h} h^{ab}, \quad (1.8)$$

and hence on-shell (classically)

$$S[g, X] \Big|_{\frac{\delta S[g, X]}{\delta g^{ab}}=0} = S_{\text{NG}}[X]. \quad (1.9)$$

The action (1.5) is quadratic in the derivatives of the 2d scalar fields $X^\mu(\sigma)$ and is expected to be better behaved at the quantum level than the non-polynomial one in Eq. (1.3), so it is natural to use (1.5) instead of the NG action for quantization. In the quantum theory, it is also natural to Wick rotate the 2d metric g_{ab} and take it to be positive-definite (i.e. to have Euclidean signature $(+, +)$). The Wick rotation makes the path integral over the 2d metric much better behaved.

The action $S[g, X]$, with Euclidean world-sheet metric g_{ab} , is the *Polyakov action* [19–21]. We look at $S[g, X]$ as defining a two-dimensional field theory living on the string world-sheet Σ , in fact as a kind of “2d gravity” since the metric is a “dynamical” field in the formal sense that we integrate over g_{ab} in the path integral. We write “dynamical” between quotes since in two dimensions the metric does not describe any local propagating d.o.f.; see below. We shall refer to this 2d QFT as the *world-sheet theory*.

We are free to add to $S[g, X]$ other terms which are *purely topological* (i.e. depend only on the topology of the world-sheet Σ), since they do not affect the classical equations of motion, nor our discussion in Eqs. (1.5)–(1.9), but are relevant in the quantum path integral. For a *connected* 2d manifold Σ , the basic topological invariant is its *Euler characteristics*¹¹

$$\chi(\Sigma) = \frac{1}{4\pi} \int_\Sigma d^2\sigma \sqrt{g} R + \frac{1}{2\pi} \int_{\partial\Sigma} ds k \equiv 2 - 2g - b, \quad (1.10)$$

¹¹ Formula (1.10) is for oriented 2-manifolds (surfaces); see BOX 1.5 for the topology of surfaces.

where R is the *scalar curvature* of g_{ab} and k the *geodesic curvature* of the curve $\partial\Sigma$:¹²

$$k = -t^a n_b D_a t^b, \quad (1.11)$$

where t^a is a unit tangent vector to $\partial\Sigma$ and n^a an outward pointing unit vector orthogonal to t^a . Equation (1.10) is the *Gauss–Bonnet theorem*: in the RHS $g \geq 0$ is the *genus* of the curve, i.e. its number of handles, and $b \geq 0$ is the number of connected components of the *boundary* $\partial\Sigma$. So, for instance:

- a sphere ($g = 0, b = 0$) has $\chi = 2$
- a disk ($g = 0, b = 1$) has $\chi = 1$
- a torus ($g = 1, b = 0$) has $\chi = 0$
- an annulus ($g = 0, b = 2$) has $\chi = 0$

We take as our starting point as the *Polyakov action*

$$S = S[g, X] + \lambda \chi(\Sigma). \quad (1.12)$$

For the moment λ is just a constant (a topological coupling), but we shall see¹³ that it has a more intrinsic interpretation as the background value of a dynamical field propagating in spacetime. Since the Euler number $\chi(\Sigma)$ depends only on the topology of Σ , the factor $e^{-\lambda\chi(\Sigma)}$ in the path integrand affects only the relative weights of different topological sectors in the sum over all world-sheets.

Adding a handle to Σ (i.e. making $g \rightsquigarrow g + 1$) reduces χ by 2, and so introduces an extra factor $e^{2\lambda}$ in the functional measure. A handle describes the process of emitting and reabsorbing a closed string (draw the cartoon!), so the amplitude for emitting a closed string is proportional to e^λ . The corresponding process for the open string introduces a new boundary component which decreases χ by 1, so the amplitude to emit an open string is proportional to $e^{\lambda/2}$. This determines the open (g_o) and closed (g_c) string coupling constants to be¹⁴

$$g_o^2 \sim g_c \sim e^\lambda. \quad (1.13)$$

Classical Symmetries

The Polyakov action has many symmetries. *Classically* $S[X, g]$ is invariant under:

(a) global target-space Poincaré invariance:

$$X^\mu \rightarrow L^\mu{}_\nu X^\nu + a^\mu, \quad g_{ab}(\sigma) \rightarrow g_{ab}(\sigma), \quad L^\mu{}_\nu \in SO(d-1, 1); \quad (1.14)$$

¹² Here and below D_a is the covariant derivative with respect to the Levi-Civita connection Γ_{bc}^a for the metric g_{ab} , and $D^2 \equiv g^{ab} D_a D_b$.

¹³ See Note 1.8.

¹⁴ Here $\sim$ means “modulo convention dependent $O(1)$ normalization coefficients”.

(b) world-sheet reparametrizations: $\sigma^a \rightarrow \sigma^{a'} = f^a(\sigma^1, \sigma^2)$

$$X^{\mu'}(\sigma') = X^\mu(\sigma), \quad g'_{ab}(\sigma') = \frac{\partial \sigma^c}{\partial \sigma'^a} \frac{\partial \sigma^d}{\partial \sigma'^b} g_{cd}(\sigma). \quad (1.15)$$

We already stated that our freedom in the choice of 2d local coordinates $\sigma \equiv (\sigma^1, \sigma^2)$ is a redundancy of the Polyakov formulation without physical meaning. In the 2d quantum field-theoretic language, the 2d reparametrizations (or, more intrinsically, the *2d diffeomorphisms* $f: \Sigma \rightarrow \Sigma$) are *gauge symmetries* (as always when the metric g_{ab} is a “dynamical” field);

(c) 2d **Weyl** invariance:

$$X^{\mu'}(\sigma) = X^\mu(\sigma), \quad g'_{ab}(\sigma) = e^{2\omega(\sigma)} g_{ab}(\sigma). \quad (1.16)$$

Going through the classical analysis in Eqs. (1.5)–(1.9), we see that two Weyl equivalent 2d metrics g_{ab} and g'_{ab} describe the *same* immersion of Σ in the physical spacetime M . So, as long as we insist that only the string world-story in spacetime, $X(\Sigma) \subset M$, has an intrinsic physical meaning, the Weyl symmetry should also be a mere redundancy of the formalism, that is, a *gauge symmetry*.

Note 1.2 To be precise: if the string is *unoriented* the full group of diffeomorphisms $\mathbf{Diff}(\Sigma)$ is a gauge symmetry. For the *oriented* string only the subgroup $\mathbf{Diff}^+(\Sigma)$ of orientation-preserving¹⁵ diffeomorphisms is a gauge symmetry: in the oriented case two string world-stories which differ in orientation are physically inequivalent.

At the quantum level some of these classical symmetries are *anomalous*. A proper treatment of the anomalies is a crucial ingredient for the correct quantization of the Polyakov action (1.12). Before going to the quantum aspects, we complete our discussion of the classical theory.

Classical Equations of Motion

For classical considerations we must revert to Minkowskian signature of the 2d metric g_{ab} . We write τ (resp. σ) for a time-like (resp. space-like) world-sheet coordinate. The equations obtained by varying the field X^μ in the Polyakov action $S[g, X]$ are

$$\partial_a(\sqrt{-g} g^{ab} \partial_b X^\mu) \equiv \sqrt{-g} D^2 X^\mu = 0, \quad (1.17)$$

i.e. the free 2d massless scalar *wave equation* in the metric g_{ab} .

The Virasoro Constraint

The variation of the action with respect to the metric defines the energy–momentum tensor T^{ab} . The standard normalization in string theory is

$$T^{ab}(\sigma) \stackrel{\text{def}}{=} -\frac{4\pi}{\sqrt{-g}} \frac{\delta S}{\delta g_{ab}(\sigma)} = -\frac{1}{\alpha'} \left(\partial^a X^\mu \partial^b X_\mu - \frac{1}{2} g^{ab} \partial_c X^\mu \partial^c X_\mu \right). \quad (1.18)$$

¹⁵ I.e. diffeomorphisms $\sigma^a \mapsto \sigma^{a'}$ whose Jacobian is everywhere positive, $\det[\partial \sigma^a / \partial \sigma'^b] > 0$.

As a consequence of **Diff**⁺ invariance, T^{ab} is conserved,

$$D_a T^{ab} = 0, \quad (1.19)$$

while the **Weyl** invariance (1.16) implies that it is *classically* traceless

$$0 = 2 \frac{\delta S}{\delta w} = g_{ab} \frac{\delta S}{\delta g_{ab}} = -\frac{1}{4\pi} \sqrt{-g} g_{ab} T^{ab}. \quad (1.20)$$

The classical equation of motion for the metric g_{ab} are then

$$T^{ab} = 0, \quad (1.21)$$

which should be understood as a *constraint* (the *Virasoro constraint*) rather than a dynamical equation.¹⁶

Boundary Conditions

If the world-sheet Σ has a boundary, $\partial\Sigma \neq \emptyset$, the equations of motion (1.17) should be supplemented by a suitable boundary condition (b.c.) on $\partial\Sigma$.

Suppose Σ is the infinite strip of width ℓ

$$\Sigma = \{(\sigma, \tau) \in \mathbb{R}^2: 0 \leq \sigma \leq \ell, -\infty < \tau < +\infty\}, \quad (1.22)$$

describing the free propagation of an open string. The value of ℓ is immaterial: it can be set to any convenient value ℓ_0 by the diffeomorphism $\sigma \mapsto (\ell_0/\ell)\sigma$. Then

$$\delta S = \frac{1}{2\pi\alpha'} \int_{-\infty}^{+\infty} d\tau \int_0^\ell d\sigma \sqrt{-g} \delta X^\mu D^2 X_\mu - \frac{1}{2\pi\alpha'} \int_{-\infty}^{+\infty} d\tau \sqrt{-g} \delta X^\mu \partial^\sigma X_\mu \Big|_{\sigma=0}^{\sigma=\ell}. \quad (1.23)$$

The vanishing of the bulk term implies the wave equation (1.17). The vanishing of the boundary term requires a suitable boundary condition. If we wish to preserve the spacetime Poincaré invariance (1.14), δX^μ cannot vanish at the boundary, and the only possibility is the *Neumann* (N) boundary condition

$$\partial^\sigma X^\mu(\tau, 0) = \partial^\sigma X^\mu(\tau, \ell) = 0, \quad (1.24)$$

¹⁶ A *constraint* is a condition which should be imposed on the *initial state*, whereas an *equation of motion* describes the evolution in time of an arbitrary initial state. A constraint is consistent with the equations of motion iff, once imposed on the initial state, it holds automatically at all later times. The Virasoro constraint says that the initial state must have vanishing energy–momentum tensor. In two dimensions, once this condition is imposed on the initial state, it remains satisfied at all times provided the energy–momentum tensor is conserved, symmetric, and *traceless*.

or, more covariantly,

$$n^a \partial_a X^\mu = 0 \quad \text{on } \partial \Sigma, \quad (1.25)$$

where n^a is a normal vector to the boundary.

The other boundary condition consistent with Poincaré invariance is that the fields are periodic on $\Sigma \equiv S^1 \times \mathbb{R}_\tau$. This corresponds to closed strings: Σ has no boundary, hence no boundary condition is required.

Alternative boundary conditions are possible (and physically meaningful): they describe configurations which are *note* invariant under the full d -dimensional Poincaré group. The most interesting such b.c. is the *Dirichlet* (D) one. Imposing the Dirichlet boundary condition on some coordinate X^{μ_0} , pointing in the direction μ_0 , at say the boundary $\sigma = 0$, means that we fix the boundary value of the 2d field X^{μ_0} rather than the value of its normal derivative

$$X^{\mu_0}(\tau, 0) = a^{\mu_0} \quad (\text{constant}). \quad (1.26)$$

Then $\delta X^{\mu_0}|_{\sigma=0} = 0$ and the boundary term in (1.23) also vanishes. At the other boundary, $\sigma = \ell$ we may impose either the N or the D boundary condition, and this can be done independently for each one of the various space directions $\mu = 1, \dots, d-1$. The deep physical implications of the Dirichlet boundary condition will be discussed in Chaps. 6 and 12.

Quantization

It remains to quantize the theory. Formally, this amounts to compute the path integral of the world-sheet 2d QFT

$$\int [dX dg] \exp(-S[X, g] - \lambda \chi)(\dots) \quad (1.27)$$

with appropriate operator insertions $(\dots)$. The path integral splits into a discrete sum over the topology classes of the world-sheet Σ and a functional integral over the continuous space of metrics g_{ab} in each given topology

$$\text{Eq. (1.27)} = \sum_{\text{top}} e^{-\lambda \chi(\text{top})} \int_{\text{top}} [dX dg] \exp(-S[X, g])(\dots). \quad (1.28)$$

However this is not quite correct: two configurations (X, g) and (X', g') related by a local **Diff**⁺ $\times$ **Weyl** symmetry are the *same* physical world-story, that is, the same oriented submanifold $X(\Sigma) \subset M$, written in two different gauges. As always in the presence of a gauge symmetry, to compute the path integral, we have first to get rid of this unphysical redundancy by fixing the gauge, and then construct the correct functional measure by the Faddeev–Popov procedure. Having done that, we may use the modern techniques of BRST quantization.¹⁷

¹⁷ BRST quantization of gauge theories was introduced in [22, 23]. For reviews see [24, 25].

Before going to the modern treatment, we shall make a brief detour into ancient history (late “pre-history”) and sketch the old-fashioned approach to string quantization which has its own merits.

1.3 Bosonic String: Light-Cone Quantization

In the old times people used the *light-cone gauge* [26]. This is a physical unitary gauge, with no ghosts, but not manifestly Lorentz covariant, akin to the *Coulomb gauge* in old-fashioned QED [27]. In the QED Coulomb gauge we get the physical Hilbert space by first solving explicitly the gauge constraint, i.e. the *Gauss law*,

$$\nabla \cdot \mathbf{E} = \varrho(\mathbf{y}, t) \quad \text{here } \varrho(\mathbf{y}, t) \text{ is the electric charge density operator} \quad (1.29)$$

in terms of physical (transverse) degrees of freedom¹⁸ [27]

$$\nabla \cdot \mathbf{A} = 0, \quad A_0(\mathbf{x}, t) = \int G(\mathbf{x}, \mathbf{y}) \varrho(\mathbf{y}, t) d^{d-1}\mathbf{y}, \quad (1.30)$$

and then performing the canonical quantization of the unconstrained transverse degrees of freedom.¹⁹ The price we pay for having a physical Hilbert space, with manifestly positive Hermitian product, is that the quantization is not manifestly Lorentz covariant since in the Coulomb gauge the splitting between “physical” and “unphysical” d.o.f. depends on a choice of Lorentz frame; cf. Eq. (1.30).

Likewise, in the string light-cone quantization one first solves explicitly the Virasoro constraint

$$T^{ab} = 0 \quad (1.31)$$

in terms of physical transverse d.o.f. and then proceeds with the canonical quantization. The main advantage of the light-cone gauge is *manifest unitarity*, i.e. its Hilbert space of states $\mathbf{H}_{\text{l.c.}}$ is positive-definite and contains only physical states (no negative-norm ghosts or unphysical longitudinal modes). From $\mathbf{H}_{\text{l.c.}}$ we can read directly the *spectrum* of the string, i.e. the list of its physical on-shell states with their quantum numbers and masses. The drawbacks of this unitary gauge are that Lorentz symmetry is not manifest (nor guaranteed) and that computations of physical processes become rather cumbersome.

Since we are mainly interested in modern covariant quantizations, we go through the light-cone one quickly: this section will be rather sketchy. The reader is invited to fill in the missing details as a good exercise.

¹⁸ $G(\mathbf{x}, \mathbf{y})$ is the $(d-1)$ dimensional massless Green function for the Laplacian which solves the equation $-D^2 G(\mathbf{x}, \mathbf{y}) = \delta^{d-1}(\mathbf{x} - \mathbf{y})$.

¹⁹ For the detailed procedure of canonical quantization of QED in the Coulomb gauge, see, for example, [27].

BOX 1.1 - Properties of 2d manifolds of Lorentzian signature

Lemma 1.1 *All 2d pseudo-Riemannian metrics of signature (1, 1) can be locally set in the form $ds^2 = -h(\sigma_+, \sigma_-)^2 d\sigma^+ d\sigma^-$ for suitable local coordinates $\sigma^\pm$.*

Proof Given a generic metric $ds^2 = g_{ij}dx^i dx^j$ of Lorentzian signature, we introduce a $SO(1, 1)$ -vielbein $e^\pm \equiv e^\pm_i dx^i$ such that

$$ds^2 = -e^+ \otimes e^-.$$

Since $SO(1, 1) \simeq \mathbb{R}$, the only non-zero component of the spin-connection is the 1-form

$$\omega^+_{+} = -\omega^-_{-} \equiv A.$$

The Levi-Civita connection is torsionless, so the first Cartan structure equation^a reduces to

$$De^\pm \equiv de^\pm \pm A \wedge e^\pm = 0$$

thus, by the Frobenius theorem [29], the two distributions defined (respectively) by the 1-forms e^+ and e^- are integrable, that is, there exist local functions σ^+ and σ^- such that

$$e^\pm_i \partial x^i / \partial \sigma^\mp = 0 \quad \Rightarrow \quad e^\pm = (e^\pm_i \partial x^i / \partial \sigma^\pm) d\sigma^\pm \equiv f^\pm d\sigma^\pm.$$

Then

$$ds^2 = -h^2 d\sigma^+ d\sigma^- \quad \text{where} \quad h^2 \equiv f^+ f^-$$

(if $f^+ f^-$ is locally negative, flip the sign $\sigma^- \rightarrow -\sigma^-$ to make it positive).

^a For the relevant Cartan structure equations, see, for example, the appendix of [28]

1.3.1 Quantization in the Light Cone

In this subsection Σ is either an infinite strip of width ℓ (cf. Eq. (1.22)) describing the free propagation of an open oriented string, or an infinite cylinder $S^1 \times \mathbb{R}$, which describes the free propagation of a closed oriented string. Consideration of these two geometries suffices to determine the perturbative spectrum of the bosonic string.

We write the light-cone coordinates in physical spacetime as

$$X^\mu \equiv (X^+, X^-, X^i), \quad ds^2 = dX^i dX^i - 2 dX^+ dX^- \quad (1.32)$$

where $X^\pm \equiv (X^0 \pm X^{d-1})/\sqrt{2}$ are the *light-cone coordinates* and X^i are the *transverse coordinates* with respect to the chosen Lorentz frame.²⁰

In physical non-covariant gauges we need to work with world-sheets of Lorentz signature. All 2d Minkowskian metrics can be set in the form

²⁰ The Latin index i takes the values $i = 1, 2, \dots, d-2$.

$$ds^2 = h(\sigma, \tau)^2 (-d\tau^2 + d\sigma^2) = -h(\sigma^+, \sigma^-)^2 d\sigma^+ d\sigma^-, \quad \sigma^\pm \equiv \tau \pm \sigma \quad (1.33)$$

by a suitable choice of local coordinates; cf. **BOX 1.1**. This choice does not fix the **Diff**⁺-gauge completely: the residual gauge transformations have locally the form

$$\sigma^\pm \rightarrow f^\pm(\sigma^\pm), \quad (1.34)$$

for arbitrary functions $f^\pm$ of their respective arguments.

In coordinates $\sigma^\pm$ the Polyakov action simplifies (we omit the topological term)

$$S = \frac{1}{4\pi\alpha'} \int d^2\sigma (\partial_\tau X^\mu \partial_\tau X_\mu - \partial_\sigma X^\mu \partial_\sigma X_\mu) \equiv \frac{1}{\pi\alpha'} \int d^2\sigma \partial_+ X^\mu \partial_- X_\mu, \quad (1.35)$$

and the wave equation (1.17) takes the simple form

$$\partial_+ \partial_- X^\mu = 0, \quad (1.36)$$

whose general solution is

$$X^\mu(\sigma^+, \sigma^-) = X_L^\mu(\sigma^+) + X_R^\mu(\sigma^-) \quad (1.37)$$

where $X_L^\mu(\sigma^+)$ and $X_R^\mu(\sigma^-)$ are *arbitrary* functions of their respective arguments. The function $X_L^\mu(\sigma^+)$ (resp. $X_R^\mu(\sigma^-)$) represents a 2d wave-packet on the world-sheet which propagates to the left (resp. to the right) at the speed of light: we call the respective d.o.f. *left-movers* and *right-movers*.

The functions $X_L^\mu(\sigma^+)$ and $X_R^\mu(\sigma^-)$ are restricted by the boundary conditions (periodic, Neumann, etc.) appropriate for the class of strings under consideration (closed oriented, open oriented, etc.). In the non-oriented case the symmetry Ω which interchanges the left- and right-movers is gauged, so the two sets of d.o.f. are not independent any longer. The same applies in the presence of a boundary; see below.

Spacetime Momentum

The spacetime momentum P^μ is the conserved 2d charge obtained by applying the Noether theorem to the global shift symmetry $X^\mu \rightarrow X^\mu + a^\mu$ of the Polyakov action:

$$P^\mu = \frac{1}{2\pi\alpha'} \int_0^\ell d\sigma \partial_\tau X^\mu. \quad (1.38)$$

We write p^μ (or k^μ) for the constant eigenvalue of the momentum operator P^μ .

Gauge-Fixing Condition

In the light-cone gauge one fixes the residual gauge freedom (1.34) by choosing the world-sheet time coordinate τ proportional to the spacetime coordinate X^+

$$X^+ = \frac{2\pi\alpha'}{\ell} p^+ \tau, \quad (1.39)$$

where the overall coefficient is fixed by Eq. (1.38). One reaches this gauge by a reparametrization of the form (1.34) with

$$f^+(\sigma^+) = \frac{\ell}{\pi \alpha' p^+} X_L^+(\sigma^+), \quad f^-(\sigma^-) = \frac{\ell}{\pi \alpha' p^+} X_R^+(\sigma^-). \quad (1.40)$$

The gauge-fixed action is (here $\dot{X}^\mu \equiv \partial_\tau X^\mu$, $\partial X^\mu \equiv \partial_\sigma X^\mu$)

$$S = \frac{1}{4\pi\alpha'} \int d\tau d\sigma \{ (\dot{X}^i)^2 - (\partial X^i)^2 \} - \int d\tau p^+ q^-, \quad (1.41)$$

where we sum over the index i labeling the *transverse* directions $i = 1, 2, \dots, d-2$, and

$$q^- \equiv \frac{1}{\ell} \int_0^\ell d\sigma X^- \quad (1.42)$$

is the string center-of-mass position in the light-cone direction X^- . p^+ is a constant of motion, so the overall coefficient in (1.39) is a constant which may be set to 1 by a convenient choice of ℓ .

Canonical Quantization

The first term in Eq. (1.41) is the 2d free action of $d-2$ massless scalars X^i with e.o.m. the wave equation (1.36). The canonical Hamiltonian of the Lagrangian (1.41)

$$H_{\text{l.c.}} = \frac{1}{2} \int_0^\ell d\sigma \left(2\pi\alpha' \Pi_i^2 + \frac{1}{2\pi\alpha'} (\partial X^i)^2 \right) \quad (1.43)$$

contains *only* the transverse degrees of freedom X^i . Here

$$\Pi_i \equiv \frac{\dot{X}^i}{2\pi\alpha'} \quad (1.44)$$

are the conjugate momenta of the 2d scalar fields X^i . The transverse fields X^i 's are the only surviving degrees of freedom in the light-cone gauge.

The Virasoro constraint takes the form

$$\partial_\pm X^- = \frac{\ell}{2\pi\alpha' p^+} (\partial_\pm X^i)^2 \quad (1.45)$$

and is explicitly solved by expressing X^- in terms of the physical degrees of freedom X^i . However some constraints still survive. First of all, Eq. (1.45) implies

$$p^- \equiv \frac{1}{2\pi\alpha'} \int d\sigma \dot{X}^- = \frac{\ell}{2\pi\alpha' p^+} H_{\text{l.c.}}. \quad (1.46)$$

In addition, in the case of *closed strings*,

$$\begin{aligned}
(\text{closed string}) \quad 0 &= \oint d\sigma \partial_\sigma X^- = \int_0^\ell d\sigma \left(\partial_+ X^- - \partial_- X^- \right) = \\
&= \frac{\ell}{2\pi\alpha'} \oint d\sigma \left((\partial_+ X^i)^2 - (\partial_- X^i)^2 \right)
\end{aligned} \tag{1.47}$$

a constraint that we need to impose on the physical closed string states.

In the light-cone approach the quantization of the string is reduced to the quantization of the free theory of the transverse scalar fields X^i . We proceed by standard canonical quantization. We consider the mode (Fourier) expansions:

- for open strings

$$X^i(\sigma, \tau) = x^i + \frac{2\pi\alpha'}{\ell} p^\mu \tau + i\sqrt{2\alpha'} \sum_{n \neq 0} \frac{1}{n} \alpha_n^i e^{-i\pi n \tau / \ell} \cos\left(\frac{n\pi\sigma}{\ell}\right) \tag{1.48}$$

- for closed strings $X^i = X_L^i + X_R^i$ with

$$X_R^i(\tau - \sigma) = \frac{1}{2}x^i + \frac{\pi\alpha'}{\ell} p^i(\tau - \sigma) + i\sqrt{\frac{\alpha'}{2}} \sum_{n \neq 0} \frac{1}{n} \tilde{\alpha}_n^i e^{-2\pi i n(\tau - \sigma)/\ell} \tag{1.49}$$

$$X_L^i(\tau + \sigma) = \frac{1}{2}x^i + \frac{\pi\alpha'}{\ell} p^i(\tau + \sigma) + i\sqrt{\frac{\alpha'}{2}} \sum_{n \neq 0} \frac{1}{n} \alpha_n^i e^{-2\pi i n(\tau + \sigma)/\ell} \tag{1.50}$$

where $\alpha_n, \tilde{\alpha}_n$ are harmonic *oscillator operators*. They satisfy the reality conditions

$$(\alpha_n^i)^\dagger = \alpha_{-n}^i \quad \text{and} \quad (\tilde{\alpha}_n^i)^\dagger = \tilde{\alpha}_{-n}^i. \tag{1.51}$$

In terms of the operators $\alpha_n^i, \tilde{\alpha}_n^i$ the canonical equal-time commutation relations

$$[X^i(\sigma, \tau), \Pi_j(\sigma', \tau)] = i\delta^i_j \delta(\sigma - \sigma') \tag{1.52}$$

become (here $n \in \mathbb{Z} \setminus \{0\}$)

$$[\alpha_n^i, \alpha_m^j] = n\delta^{ij} \delta_{m+n,0}, \quad [\tilde{\alpha}_n^i, \tilde{\alpha}_m^j] = n\delta^{ij} \delta_{m+n,0}, \tag{1.53}$$

that is, we have one harmonic oscillator of frequency

$$\omega_n = n/\ell \tag{1.54}$$

per each index $i = 1, \dots, d-2$ and positive integer $n > 0$: α_n^i are annihilator operators and their Hermitian conjugates α_{-n}^i are creator operators (the same story goes for the tilded operators $\tilde{\alpha}_n^i$). Hence

- for the closed string

$$H_{\text{l.c.}} = \frac{\pi\alpha'}{\ell} p^i p^i + \frac{2\pi}{\ell} \left(\sum_{n>0} (\alpha_{-n}^i \alpha_n^i + \tilde{\alpha}_{-n}^i \tilde{\alpha}_n^i) + a + \tilde{a} \right) \quad (1.55)$$

- for the open string

$$H_{\text{l.c.}} = \frac{\pi\alpha'}{\ell} p^i p^i + \frac{\pi}{\ell} \left(\sum_{n>0} \alpha_{-n}^i \alpha_n^i + a \right) \quad (1.56)$$

with $a, \tilde{a}$ constants called the *shifts*. In the old times the additive constant $a \equiv \tilde{a}$ was interpreted as due to a “normal-order ambiguity” in the product of creation/annihilation operators—in this textbook we shall give (several) more intrinsic interpretations of this constant when discussing the modern versions of string quantization. The simplest (and older) interpretation of the quantity $2\pi a/\ell$ is as the zero-point energy of the harmonic oscillators; indeed

$$a = \frac{\ell}{2\pi} \cdot \frac{1}{2} \sum_{i,n} \omega_n = \frac{d-2}{2} \sum_{n>0} n \xrightarrow{\zeta\text{-reg.}} \frac{d-2}{2} \zeta(-1) \equiv -\frac{(d-2)}{24} \quad (1.57)$$

where we used ζ -function regularization [30, 31]. Equivalently, $a + \tilde{a} \equiv 2a$ is the Casimir energy of $(d-2)$ free massless scalars on a circle S^1 of length $\ell = 2\pi$ with periodic boundary conditions. Formulae for the Casimir energies of 2d free fields with various boundary conditions are important in string theory; the story is summarized in **BOX 1.2**.

The constraint (1.46) becomes the physical on-shell mass condition:

$$\text{(closed string)} \quad m^2 \equiv 2 p^+ p^- - p^i p^i = \frac{2}{\alpha'} (N_{\text{tr}} + \tilde{N}_{\text{tr}} + a + \tilde{a}) \quad (1.58)$$

$$\text{(open string)} \quad m^2 \equiv 2 p^+ p^- - p^i p^i = \frac{1}{\alpha'} (N_{\text{tr}} + a), \quad (1.59)$$

where the *transverse oscillator numbers* (left and right) are

$$N_{\text{tr}} = \sum_{n>0} \alpha_{-n}^i \alpha_n^i, \quad \tilde{N}_{\text{tr}} = \sum_{n>0} \tilde{\alpha}_{-n}^i \tilde{\alpha}_n^i. \quad (1.60)$$

The open string Hilbert space contains an *infinite tower* of states of the form

$$\alpha_{-n_1}^{i_1} \alpha_{-n_2}^{i_2} \cdots \alpha_{-n_s}^{i_s} |0; p_\mu\rangle, \quad (1.61)$$

BOX 1.2 - Casimir energies for free massless fields on S^1

We set the length of the circle S^1 to $\ell = 2\pi$ without loss of generality. The simplest way to compute the Casimir energies a is to use the ζ -function regularization [30, 31] for the zero-point energy, which for free bosons (resp. fermions) is one-half (resp. minus one-half) the sum of the frequencies of all field oscillators

$$a = \pm \frac{1}{2} \sum_{n \in \mathbb{N}} \omega_n.$$

We take the free field ϕ to be complex with the generalized periodic boundary condition

$$\phi(w + 2\pi) = e^{2\pi i x} \phi(w) \quad (0 \leq x < 1).$$

Since a complex field corresponds to 2 real ones, this gives an extra factor of 2; therefore if the fields are real (possible only for $2x \in \mathbb{Z}$), the following expressions must be divided by 2. We focus on the contribution a to the zero-point energy from the left-movers. The contribution $\tilde{a}$ from right-movers takes the same value. Since $\omega_n = n - x$, we have

$$\begin{aligned} a(x)_{\text{bos}} &= -a(x)_{\text{fer}} = \tilde{a}(1-x)_{\text{bos}} = -\tilde{a}(1-x)_{\text{fer}} = \\ &= \sum_{n \geq 1} (n-x) = \sum_{n \geq 1} \frac{1}{(n-x)^s} \Big|_{s=-1} \equiv \zeta(s, 1-x) \Big|_{s=-1} \end{aligned}$$

where $\zeta(s, z)$ is the *Hurwitz ζ -function* (§.25.11 of [32]). While the sum $\sum_{n \geq 0} (n+z)^{-s}$ converges only for $\text{Re } s > 1$, the function $\zeta(s, z)$ itself exists as a meromorphic function in $\mathbb{C}$ whose only singularity is a simple pole at $s = 1$. For s a negative integer, $\zeta(-n, z)$ is equal to $-B_{n+1}(z)/(n+1)$ with $B_k(z)$ the (periodic) *k-th Bernoulli polynomial* (§.24 of [32]). One has

$$\zeta(-1, z) = -\frac{1}{2} B_2(z) = -\frac{1}{2} z^2 + \frac{1}{2} z - \frac{1}{12}.$$

Here $B_2(z)$ is the second Bernoulli polynomial. We conclude that

$$a(x)_{\text{bos}} = -\frac{1}{12} + \frac{x(1-x)}{2}.$$

For *real* periodic fields ($x = 0$) this yields $\mp 1/24$ while for *real* anti-periodic ones ($x = 1/2$), this gives $\pm 1/48$ the upper (lower) sign being for bosons (resp. fermions).

where $|0; p_\mu\rangle$ is the (transverse oscillators' boosted) vacuum which satisfies²¹

$$\alpha_n^i |0; p_\mu\rangle = \tilde{\alpha}_n^i |0; p_\mu\rangle = 0 \quad \text{for } n > 0, \quad P_\mu |0; p_\mu\rangle = p_\mu |0; p_\mu\rangle. \quad (1.62)$$

The infinitely many states of the form (1.61) are all *normalizable* eigenstates of $m^2 \equiv -P^\mu P_\mu$, hence “particles” propagating in the physical d -dimensional spacetime. Of course most of them will decay as soon as we take into account string interactions.

The transverse oscillator number operators $N_{\text{tr}}, \tilde{N}_{\text{tr}}$ are Hermitian and non-negative with integral spectrum:

²¹ In Chap. 2 we shall give a more intrinsic interpretation/derivation of these equations.

$$N_{\text{tr}} \alpha_{-n_1}^{i_1} \alpha_{-n_2}^{i_2} \dots \alpha_{-n_s}^{i_s} |0; p_\mu\rangle = \left(\sum_{k=1}^s n_k \right) \alpha_{-n_1}^{i_1} \alpha_{-n_2}^{i_2} \dots \alpha_{-n_s}^{i_s} |0; p_\mu\rangle. \quad (1.63)$$

A state $|\psi\rangle$ is said to have (left) *level* n iff

$$N_{\text{tr}} |\psi\rangle = n |\psi\rangle. \quad (1.64)$$

States of zero level are also called *ground states*. *Closed string* states have the form

$$\tilde{\alpha}_{-m_1}^{j_1} \tilde{\alpha}_{-m_2}^{j_2} \dots \tilde{\alpha}_{-m_r}^{j_r} \alpha_{-n_1}^{i_1} \alpha_{-n_2}^{i_2} \dots \alpha_{-n_s}^{i_s} |0; p_\mu\rangle \quad (1.65)$$

subjected to the mass condition (1.58) and the surviving constraint (1.47)

$$0 = \frac{2}{\ell\alpha'} \int d\sigma \left(\partial_+ X^i \partial_+ X^i - \partial_- X^i \partial_- X^i \right) = N_{\text{tr}} - \tilde{N}_{\text{tr}}, \quad (1.66)$$

i.e. we must impose equality of left and right (transverse) oscillation numbers:

$$\sum_i n_i = \sum_j \tilde{n}_j. \quad (1.67)$$

This constraint is called the *left–right matching condition*.

1.3.2 Lorentz Invariance: Emergence of Gravity

The main drawback of the light-cone approach is that it is not clear whether the theory is Lorentz covariant in the target-space M . Indeed, for generic spacetime dimension d (and shift constant a) Lorentz symmetry is *anomalous*. The simplest way to see this is to consider the first excited states ($N_{\text{tr}} = 1$), say in the open string

$$\alpha_{-1}^i |0; p_\mu\rangle_{\text{open}} \quad i = 1, 2, \dots, d-2. \quad (1.68)$$

These states form a *single* vector of the $SO(d-2)$ which rotates the transverse directions X^i . In all Lorentz invariant theory, massive particles form representations of the spin group $SO(d-1)$ and massless ones representations of the $SO(d-2)$ little group of the null vector p^μ . Hence the string spectrum is consistent with Lorentz symmetry only if the mass of the states (1.68) is zero. From Eqs. (1.59) and (1.57)

$$m^2 = \frac{1}{\alpha'} \left(1 - \frac{d-2}{24} \right) \quad (1.69)$$

so, at the very best, we have Lorentz invariance only when the target space has dimension $d = 26$ which is called the *critical dimension* for the bosonic string.

Lorentz invariance for $d = 26$ comes with a heavy price: the open string ground state $|0; p_\mu\rangle_{\text{open}}$ is a *tachyonic scalar* with mass-squared

$$m^2 = -1/\alpha'. \quad (1.70)$$

The presence of a tachyon signals a *quantum instability* of the naive vacuum in which we are considering bosonic string theory.

Emergence of Gravity

Due to the left–right matching condition (1.66), the first excited states in the closed string sector are

$$\alpha_{-1}^i \tilde{\alpha}_{-1}^j |0; p_\mu\rangle_{\text{closed}}. \quad (1.71)$$

In critical dimension $d = 26$ they are massless states transforming in three distinct irreducible representations of the transverse Lorentz group $SO(24)$:

- the trivial representation (trace part)
- the symmetric traceless part
- the antisymmetric part.

They correspond to three massless particles propagating in physical spacetime which are described, respectively, by a scalar field Φ (the *dilaton*), a symmetric 2-tensor $G_{\mu\nu}$ (the *spacetime metric*), and a 2-form field $B_{\mu\nu}$ (the *B-field*).

We know that a (non-free) massless spin-2 field $G_{\mu\nu}$ is consistent if and only if the theory contains dynamical gravity [15–18]. We conclude that **a non-trivial, consistent, Lorentz covariant, quantum string theory describes inter alia dynamical gravity in physical spacetime**. We shall check later that string theory does satisfy the gravitational Ward identities, and so is a *bona fide* Quantum Gravity. The most important problem in theoretical physics is to formulate a fully consistent quantum theory of dynamical gravity. String theory is on the right track to do that, and this is the ultimate reason why we are interested in it.

Exercise 1.1 Show that for $d = 26$ the states of the first massive level ($N_{\text{tr}} = 2$) form representations of $SO(d - 1)$ as they should.

More generally, for (say) the open string, let

$$\mathbf{H}_{\text{open string light-cone}} = \bigoplus_{n \geq 0} \mathbf{H}_n \quad (1.72)$$

be the decomposition of the physical Hilbert space into subspaces of definite level n . Lorentz invariance requires $\mathbf{H}_1$ to carry a representation of $SO(24)$ while for all $n \geq 2$, $\mathbf{H}_n$ should carry a natural representation of $SO(25)$. To check these statements for all n is a very hard exercise in combinatorics. Luckily, there is a handy trick which makes matters easy. The trick comes from an unexpected venue: Algebraic Geometry; see **BOX 1.3**.

BOX 1.3 - Lorentz symmetry, critical dimension, and K3 surfaces

Let $\mathbf{H}_n$ be the level- n subspace in Eq. (1.72). We write $\mathbf{H}_n(p) \subset \mathbf{H}_n$ for the eigen-subspace with a fixed on-shell momentum p^μ (with $p^\mu p_\mu = -(n-1)/\alpha'$).

In Chap. 12, when discussing the *non-perturbative* aspects of string theory, we shall encounter the following theorems of Algebraic Geometry:

Theorem (Göttsche [33]). *There exists a canonical isomorphism*

$$\mathbf{H}_n(p) \cong H^\bullet(K3^{[n]}, \mathbb{C}) \equiv \bigoplus_k H^{2k}(K3^{[n]}, \mathbb{C})$$

between the eigen-subspace $\mathbf{H}_n(p) \subset \mathbf{H}_n$ and the total cohomology space of the Hilbert scheme of n -points, $K3^{[n]}$, on an algebraic K3 surface.^a

Theorem (Beauville [34]). $K3^{[n]}$ is a compact, simply connected, hyperKähler manifold of complex dimension $2n$. ($K3^{[1]}$ is just the surface $K3$). Their second Betti numbers are

$$b_2(K3^{[n]}) \equiv \dim H^2(K3^{[n]}, \mathbb{C}) = \begin{cases} 22 & n = 1 \\ 23 & n \geq 2. \end{cases}$$

In particular $\chi(K3) = 24$.

Theorem (Verbitsky [35]). *On the total real cohomology of a compact hyperKähler manifold, there is a canonical action of the Lie algebra $\mathfrak{so}(4, b_2 - 2)$.*

Corollary. *On the complex spaces $\mathbf{H}_n$, there is a canonical Verbitsky action of the group $SO(24)$ for $n = 1$ and of the group $SO(25)$ for $n \geq 2$.*

Then it appears that the critical dimension d of a Lorentz-covariant string theory should be

$$d = 2 + \sum_{k \geq 0} b_{2k}(S)$$

for some compact hyperKähler surface S . There are *two* compact hyperKähler surfaces: $K3$ and the 4-torus T^4 . We get two possible critical dimensions: 26 and 10.

Suggestion: there must be a non-trivial string theory which lives in 10 dimensions!

^a “Surface” is used in the complex sense. It stands for “compact complex manifold of complex dimension 2”.

1.4 Covariant Quantization *à la* Polyakov

Following Polyakov [19, 36], we compute the path integral (1.27) by fixing a convenient *covariant* gauge. The path integral is taken over 2d metrics g_{ab} of Euclidean ($\equiv$ positive) signature. Given a two-dimensional Riemannian manifold Σ , we may always find *local* coordinates σ^1, σ^2 such that the metric locally takes the form

$$ds^2 = e^{2\omega(\sigma)} ((d\sigma^1)^2 + (d\sigma^2)^2). \quad (1.73)$$

This is the Euclidean-signature version of the Minkowskian **Lemma** proven in **BOX 1.1**. A different (and more convenient) perspective on the same result is given in **BOX 1.4**. For reasons better explained in that **BOX**, it is convenient to introduce the *local* complex coordinate $z = \sigma^1 + i\sigma^2$. After fixing part of the $\mathbf{Diff}^+$ invariance by the choice of coordinates (1.73), the only surviving degree of freedom in the 2d metric is the (local) real function $\omega(z, \bar{z})$.

1.4.1 World-Sheet Topologies. Non-orientable Σ 's

The string partition function is a sum over the allowed topologies of the world-sheet Σ ; cf. Eq. (1.28). If the string is oriented, only oriented world-sheets enter in the sum, otherwise we need to sum also over non-orientable topologies. Besides, if the strings are open we must allow for world-sheets with boundaries, and we must sum over the number b of connected boundary components. It is important to understand the topological classification of 2d manifolds in order to know exactly *over what* we have to sum. The basic facts about the topological classification of closed 2-manifolds are summarized in **BOX 1.5** for the reader's convenience.

In **BOX 1.4** it is shown that all non-orientable Riemannian 2-manifold Σ has an *oriented double* $\check{\Sigma}$ whose metric is Kähler (with respect to *some* complex structure) with an anti-holomorphic isometry $\Omega: \check{\Sigma} \rightarrow \check{\Sigma}$ with $\Omega^2 = \text{Id}$ and no fixed points, such that

$$\Sigma = \check{\Sigma} / \langle \Omega \rangle. \quad (1.74)$$

The amplitude on the non-orientable world-sheet Σ is then obtained from the one on $\check{\Sigma}$ by gauging the $\mathbb{Z}_2$ symmetry generated by Ω .

We conclude that (replacing Σ with its oriented double if necessary) we may assume the world-sheet to be a Kähler manifold of complex dimension 1. From Eq. (1.73) it is obvious that a conformal class of 2d Kähler metrics is the same as a *complex structure* on Σ (cf. **BOX 1.4**). A complex manifold of complex dimension 1 is traditionally called a *Riemann surface*, and a conformal class of oriented Riemannian 2-manifolds is the same as a Riemann surface.

We shall mainly concern ourselves with *oriented closed strings*, the extension to other kinds of strings being totally straightforward. In this textbook the world-sheet Σ is assumed to be *oriented* unless explicitly stated otherwise.

1.4.2 Conformal Killing Vectors and Complex Automorphisms

The condition (1.73) does not fix completely the $\mathbf{Diff}^+$ -gauge. Recall that an infinitesimal diffeomorphism is an infinitesimal *conformal motion* if it is generated by a vector field v such that²²

²² The symbols $\mathcal{L}_v$ stands for the Lie derivative along the vector field v ; see [41, 42].

BOX 1.4 - Local and global structure of 2d Riemannian manifolds

Local structure

Lemma 1.2 Locally, all 2d Riemannian metrics can be set in the form

$$ds^2 = e^{2\omega(\sigma)} (d\sigma^1 + i d\sigma^2)(d\sigma^1 - i d\sigma^2) \quad \clubsuit$$

for suitable local coordinates σ^1, σ^2 .

Proof By definition, the holonomy Lie algebra^a of any Riemannian manifold of dimension n is contained in $\mathfrak{so}(n)$. In the case $n = 2$ one has the isomorphism

$$\mathfrak{so}(2) \simeq \mathfrak{u}(1) \equiv \mathfrak{u}(n/2),$$

hence the metric ds^2 —when pulled back to the universal cover $\tilde{\Sigma}$ of our 2d manifold Σ —has Riemannian holonomy $U(1) \equiv \exp \mathfrak{u}(1)$ —and so it is Kähler, hence complex and Hermitian. Indeed, the Levi-Civita connection is a torsionless connection with values in $\mathfrak{u}(n/2) \subset \mathfrak{gl}(n/2, \mathbb{C})$ and hence defines an *integrable* complex structure on $\tilde{\Sigma}$ by the Newlander–Nirenberg theorem. Thus, *locally*, we may always find a complex coordinate z such that

$$ds^2 = e^{2\omega(z, \bar{z})} dz d\bar{z}$$

for some real function $\omega(z, \bar{z})$. To get eq.($\clubsuit$) set $\sigma^1 = \operatorname{Re} z$, $\sigma^2 = \operatorname{Im} z$. □

Global structure

Globally there are two possibilities: our 2d manifold

$$\Sigma \equiv \tilde{\Sigma}/\Pi, \quad \Pi \subset \mathbf{Diff}(\tilde{\Sigma}), \quad \Pi \simeq \pi_1(\Sigma) \text{ is the universal deck group}$$

may be *orientable* or *non-orientable*, since Π may be a group of diffeomorphisms which may or may not preserve the canonical orientation $i dz \wedge d\bar{z}$ of $\tilde{\Sigma}$. In the orientable case the 2-manifold Σ is automatically Kähler.

Non-orientable case: the oriented double

In the non-orientable case, let

$$\check{\Pi} \equiv \Pi \cap \mathbf{Diff}^+(\tilde{\Sigma}) \triangleleft \Pi$$

be the subgroup preserving the orientation. $\check{\Pi}$ fits in an exact sequence of group homomorphisms

$$1 \rightarrow \check{\Pi} \rightarrow \Pi \rightarrow \mathbb{Z}_2 \rightarrow 1.$$

^a For Riemann holonomy, Kähler manifolds, and Newlander–Nirenberg theorem, see Sect. 11.1.1; a more detailed discussion for physicists in Chap. 3 of the book [37].

Then $\tilde{\Sigma} \equiv \tilde{\Sigma}/\tilde{\Pi}$ is an *oriented double cover* of the non-orientable 2d manifold $\Sigma \equiv \tilde{\Sigma}/\Pi$ called its *orientable double* (unique up to isomorphism). The orientable double is automatically Kähler.

Note 1.3 It follows from the above **Lemma** that, for an oriented 2d manifold Σ , a *conformal structure* is the same thing as a *complex structure*.

Theorem 1.1 A compact complex manifold of dimension 1 is (the analytic space underlying) a normal projective algebraic curve over $\mathbb{C}$, i.e. it is the zero locus of a finite-family of homogenous polynomials in the projective space $\mathbb{P}^N(\mathbb{C})$ for some N . Conversely all normal projective curves over $\mathbb{C}$ are compact complex manifolds.

Proof Immediate consequence of Kodaira embedding theorem; see, for example, §. 1.4 of [38].

Note 1.4 Specifying a conformal structure on an orientable compact manifold of *real* dimension 2 is then equivalent to specifying an algebraic structure over $\mathbb{C}$.

BOX 1.5 - Topological classification of closed 2-manifolds

Closed 2-manifolds are either *orientable* or *non-orientable*. A non-orientable surface has an *orientable double* (cf. **BOX 1.4**), so it has the form $S/\langle\Omega\rangle$ where S is an orientable surface and $\Omega: S \rightarrow S$ with $\Omega^2 = \text{Id}$ which acts freely on S . The simplest example of a closed orientable surface is the sphere S^2 . Another simple example is the 2-torus $T^2 \equiv S^1 \times S^1$. The simplest example of a closed non-orientable surface is the real projective plane $\mathbb{RP}^2$ (called the *cross-cap* in the string literature) which is the quotient of S^2 under the antipodal map: $z \mapsto -1/\bar{z}$ in the complex notation.

The basic topological operation is the *connected sum* of two topological surfaces: $\Sigma_1 \# \Sigma_2$ is the topological surface obtained by cutting small disks on the two surfaces Σ_1, Σ_2 and gluing the resulting boundaries together. One has $\Sigma \# S^2 = \Sigma$ for all Σ . We quote the basic result without proofs (they can be found in [39, 40]).

Theorem 1.2 All closed orientable surfaces are the connected sum of $g \geq 0$ 2-tori

$$F_g = \#^g T^2 \equiv T^2 \# T^2 \# \dots \# T^2 \quad g \text{ copies of } T^2.$$

All closed non-orientable surfaces are the connected sum of $h \geq 1$ real projective planes

$$R_h = \#^h \mathbb{RP}^2 \equiv \mathbb{RP}^2 \# \mathbb{RP}^2 \# \dots \# \mathbb{RP}^2 \quad h \text{ copies of } \mathbb{RP}^2.$$

More precisely, the set of homeomorphism classes of surfaces is a commutative monoid with respect to the connected sum, generated by T^2 and $\mathbb{RP}^2$, with the sole relation

$$T^2 \# \mathbb{RP}^2 = \mathbb{RP}^2 \# \mathbb{RP}^2 \# \mathbb{RP}^2.$$

We call g the *genus* of the surface and h the *number of cross-caps*.

Theorem 1.3 *The integral homology is*

$$\begin{aligned} H_0(F_g, \mathbb{Z}) &\cong \mathbb{Z}, & H_1(F_g, \mathbb{Z}) &\cong \mathbb{Z}^{2g} & H_2(F_g, \mathbb{Z}) &\cong \mathbb{Z} \\ H_0(R_h, \mathbb{Z}) &\cong \mathbb{Z}, & H_1(R_h, \mathbb{Z}) &\cong \mathbb{Z}^{h-1} \oplus \mathbb{Z}/2\mathbb{Z} & H_2(R_h, \mathbb{Z}) &\cong 0 \end{aligned}$$

and hence the Euler characteristic is

$$\chi\left(\left(\#^g T^2\right) \# \left(\#^h \mathbb{R}P^2\right)\right) = 2 - 2g - h.$$

Theorem 1.4 *The oriented double of R_h is F_{h-1} . In particular*

$$\pi_n(R_h) = \pi_n(F_{h-1}) \quad \text{for } n \geq 2.$$

Note 1.5 The process of taking the connected sum of a surface Σ with h cross-caps $\mathbb{R}P^2$ is visualized as follows: cut h small disks out of Σ getting a surface Σ_h with h boundary components, each of them a copy of S^1 . Parametrize each boundary component by an angle $\theta \in [0, 2\pi]$ and then identify the points on each boundary by $\theta \sim \theta + \pi$, with the effect of closing up the surface. One also says that we have h cross-caps connected by the (open) surface Σ_h .

Example: the Klein bottle

The *Klein bottle* is the non-oriented surface R_2 . Its orientable double is the 2-torus F_1 . Model the 2-torus as the quotient of $\mathbb{C}$ with coordinate $z = x + iy$ by the lattice Λ generated by 2π and $2i\pi t$ ($t \in \mathbb{R}$), $T^2 \equiv \mathbb{C}/\Lambda$. We take $\Omega: T_2 \rightarrow T_2$ to be $\Omega: z \mapsto \bar{z} + \pi$. As a fundamental domain we may consider the rectangle $0 \leq x \leq 2\pi, 0 \leq y \leq t$ in which the two vertical sides are identified while the two horizontal sides are circles parametrized by $0 \leq x \leq 2\pi$ with the identification $x \sim x + \pi$, that is, two cross-caps: *the quotient of the torus by the orientation-reversing involution Ω is then two cross-caps connected by a cylinder, that is, $\mathbb{R}P^2 \# \mathbb{R}P^2$, i.e. the Klein bottle.* We can get a different (and more common) presentation of the Klein bottle by using as fundamental domain the rectangle $0 \leq x' \leq \pi, -t \leq y' \leq t$. The horizontal sides are identified making a cylinder. The vertical sides then become the two circular boundaries of the cylinder which get identified with an inversion of the orientation $y'|_{x'=0} = -y'|_{x'=\pi}$.

$$\mathbb{L}_v g_{ab} \equiv D_a v_b + D_b v_a = 2\omega(\sigma) g_{ab}, \quad (1.75)$$

for some function $\omega(\sigma)$: we say that such a v is a *conformal Killing vector* (CKV). In other words, the conformal motions are diffeomorphisms whose effect on the 2d metric is a Weyl rescaling, i.e. they belong to the intersection

$$(\mathbf{Diff}^+) \cap (\mathbf{Weyl}). \quad (1.76)$$

Clearly the condition (1.73) fixes the $\mathbf{Diff}^+$ symmetry only up to *local* conformal motions, that is, conformal motions which act independently on the local coordinates of each coordinate chart. *Local* conformal motions (which are mere redefinitions of the local coordinates) should not be confused with *global* conformal motions to be discussed momentarily.

Table 1.1 $\text{Aut}(\Sigma)$ for some orientable 2d manifold Σ with $\partial\Sigma = \emptyset$

genus g	$\text{Aut}(\Sigma)$	$\dim_{\mathbb{C}} \text{Aut}(\Sigma)$	Notes
$g = 0$ (sphere)	$PSL(2, \mathbb{C})$	3	Projective motions of $\mathbb{P}^1$
$g = 1$ (torus)	$E \rtimes \text{Aut}(E)$	1	E torus as an Abelian group, $\text{Aut}(E) = \mathbb{Z}_2, \mathbb{Z}_4, \mathbb{Z}_6$
$g > 1$	finite	0	$\# \text{Aut}(\Sigma) \leq 84(g-1)$

Table 1.2 $\text{Aut}(\Sigma)$ for some orientable 2d manifold Σ with $\partial\Sigma \neq \emptyset$

Euler number χ	$\text{Aut}(\Sigma)$	$\dim_{\mathbb{R}} \text{Aut}(\Sigma)$
$\chi = 1$ (disk)	$PSL(2, \mathbb{R})$	3
$\chi = 0$ (cylinder)	$U(1) \equiv S^1$	1
$\chi < 0$	Finite	0

Exploiting the fact that the metric is *locally* Kähler²³ (in the sense of **BOX 1.4**), in the local complex coordinate z Eq. (1.75) reduces to

$$\partial_{\bar{z}} v^z = 0, \quad (1.77)$$

i.e. *local* conformal motions are just *holomorphic* changes of the local coordinate

$$z \rightsquigarrow z' = f(z) \quad \Rightarrow \quad ds^2 = e^{2(\omega - \log |\partial_z f|)} dz' d\bar{z}'. \quad (1.78)$$

In the new coordinate z' , the metric is still of the form (1.73) with $\omega \rightsquigarrow \omega - \log |\partial_z f|$.

If Σ is *oriented*, its *global* conformal motions form the Lie group²⁴ of its complex automorphisms $\text{Aut}(\Sigma)$. See Tables 1.1 and 1.2 for examples of automorphism groups. The automorphism group the sphere S^2 is further studied in **BOX 1.11**.

1.4.3 The Fadeev–Popov Determinant

The 2d metric has no *local* $\text{Diff}^+ \times \text{Weyl}$ gauge-invariant content.²⁵ At the *global* level the gauge-invariant content of the 2d metric g_{ab} is its conformal class—a datum

²³ Hence (by definition of “Kähler”; cf. Chap. 11) the Levi-Civita connection D coincides with the (unique) Chern connection [38, 43, 44] on the holomorphic line bundle $T\Sigma$ of $(1,0)$ vector fields $v^z \partial_z$. For a background on holomorphic line bundles, see **BOX 1.6**; for further material, see [45].

²⁴ For a proof that the complex automorphisms form a Lie group, see [46].

²⁵ Indeed, *locally* the metric can always be set in the universal form $ds^2 = dz d\bar{z}$ by a change of coordinates followed by a Weyl rescaling.

which is equivalent to specifying a *complex structure* on Σ ; cf. Note 1.4 in **BOX 1.4**. The inequivalent complex structures on Σ are parametrized by a *finite-dimensional* space $\mathcal{M}$, called the (*complex*) *moduli space* to be discussed in Sect. 1.6 below.

Up to diffeomorphisms, all 2d metric may be written in the form

$$g_{ab} = e^{2\phi} h_{ab}(m), \quad (1.79)$$

where $h_{ab}(m)$ is a *chosen* reference metric in its conformal class²⁶ specified by the point $m \in \mathcal{M}$ which (by construction) is Kähler in complex structure m . We may think of the function ϕ as a global 2d scalar field living on Σ . The gauge condition

$$g_{ab} = h_{ab}(m) \quad (1.80)$$

fixes *most* of the $\mathbf{Diff}^+ \times \mathbf{Weyl}$ gauge symmetry. We shall dwell on the residual gauge symmetry momentarily. Under $\mathbf{Diff}^+ \times \mathbf{Weyl}$ infinitesimal transformations of parameters (v^a, Λ) , the metric transforms as

$$\delta g_{ab} = (D_a v_b + D_b v_a) + 2\Lambda g_{ab} = (Pv)_{ab} + 2\tilde{\Lambda} g_{ab}, \quad (1.81)$$

where P is a first-order differential operator which maps vector fields into traceless symmetric tensors

$$(Pv)_{ab} \stackrel{\text{def}}{=} D_a v_b + D_b v_a - (D_c v^c) g_{ab}, \quad \tilde{\Lambda} \equiv \Lambda + \frac{1}{2} D_c v^c. \quad (1.82)$$

Since all *local* d.o.f. in g_{ab} are gauge ones, the Faddeed–Popov determinant Δ_{FP} is *formally* the functional Jacobian $\mathcal{J}$ between the functional measure expressed in terms of the metric g_{ab} and the invariant measure on the “gauge group” $\mathbf{Diff}^+ \times \mathbf{Weyl}$

$$[dg_{ab}] = \mathcal{J}[h] [dv d\Lambda], \quad \mathcal{J}[h] \equiv \Delta_{\text{FP}}[h], \quad (1.83)$$

where $[dv d\Lambda]$ is the (formal) volume form on the gauge “group” $\mathbf{Diff}^+ \times \mathbf{Weyl}$. We decompose g_{ab} into the traceless and trace parts writing $g_{ab} = s_{ab} + 2t h_{ab}$. Then

$$\mathcal{J}[h] = \det \begin{bmatrix} \frac{\delta s_{ab}}{\delta v^c} & 0 \\ \frac{\delta t}{\delta v^c} & \frac{\delta t}{\delta \Lambda} \end{bmatrix} = \det \begin{bmatrix} P & 0 \\ \frac{1}{2} \nabla^\dagger & 1 \end{bmatrix} = \det P, \quad (1.84)$$

where—for the moment—we ignore a crucial issue, namely the possibility that P or $P^\dagger$ have *zero-modes*. We shall discuss this fundamental aspect momentarily; for now let us close our eyes and proceed *as if* there were no zero-mode issues.

²⁶ For example, if Σ is closed we may take as $h_{ab}(m)$ the *unique* metric on Σ with constant scalar curvature $R = \text{const.}$ and unit volume which is Kähler in complex structure $m \in \mathcal{M}$.

As always, we can represent a functional determinant as a Gaussian integral over Fermi fields—the Faddeev–Popov (FP) ghosts and anti-ghosts [47]—

$$\det P = \int [db\,dc] \exp\left(-\frac{1}{2\pi} \int_{\Sigma} d^2\sigma \sqrt{g} g^{ab} b_{bc} D_a c^c\right), \quad (1.85)$$

where c^a is a complex fermion which transforms as a 2d vector, while b_{bc} is a complex fermion which is a 2d traceless symmetric tensor. Equation (1.85) agrees with the standard rule for Faddeev–Popov fields: the ghosts are complex fermions carrying the quantum numbers of the gauge parameters—for **Diff** they are vector fields v^a —while the anti-ghosts are complex fermions with the quantum numbers of the gauge fixing condition, which in our case reads $(g_{ab})_{\text{traceless}} = 0$. We write $b, \tilde{b}$ for the non-zero components $b_{zz}, b_{\bar{z}\bar{z}}$ of the anti-ghost, and $c, \tilde{c}$ for $c^z, c^{\bar{z}}$ (respectively). Then²⁷

$$\det P = \int [db\,dc\,d\tilde{b}\,d\tilde{c}] \exp\left(-\frac{1}{2\pi} \int_{\Sigma} (b\bar{\partial}c + \tilde{b}\partial\tilde{c})\right). \quad (1.86)$$

Under holomorphic reparametrizations ($\equiv$ local conformal motions) of the form

$$z \rightarrow z' \equiv f(z) \quad (1.87)$$

these ghosts transform as vectors and quadratic differentials, respectively,

$$c'(z') \partial_{z'} = c(z) \partial_z \quad \text{i.e.} \quad c'(z') = c(z)(\partial z'/\partial z) = f' c(z) \quad (1.88)$$

$$b'(z') (dz')^2 = b(z) (dz)^2 \quad \text{i.e.} \quad b'(z') = b(z)(\partial z/\partial z')^2 = (f')^{-2} b(z) \quad (1.89)$$

here $f' \equiv \partial f/\partial z$. These equations show that the ghost action (1.86) is *classically* invariant under conformal transformations of Σ .

Exercise 1.2 When the world-sheet Σ has a boundary (open string case), we need a boundary condition on the ghost fields $b(z), c(z)$. Determine the correct one.

1.4.4 The Matter Sector

Classically the Weyl factor $e^{2\phi}$ in the metric (1.79) drops out from the action which depends only on the Weyl invariant combination $\sqrt{g} g^{ab}$: in complex coordinates the matter part of the classical action is simply

$$S_{\text{matter}} = \frac{1}{\pi\alpha'} \int \partial X^\mu \bar{\partial} X_\mu. \quad (1.90)$$

²⁷ Here $\partial \equiv dz \partial_z$ and $\bar{\partial} \equiv d\bar{z} \partial_{\bar{z}}$ are Dolbeault differential operators [38, 44]. The integrands in Eqs. (1.86), (1.90) are differential forms of type (1, 1) which may be integrated over a (compact) complex manifold of complex dimension 1.

However, while the classical action contains the 2d metric only in the Weyl invariant combination $\sqrt{g}g^{ab}$, the full metric g_{ab} enters in the definition of the path integral *measure*, and the Weyl symmetry may be anomalous at the quantum level. One needs to compute the quantum amplitudes, and then check a posteriori whether they are independent of the conformal factor $e^{2\phi}$; we shall do that in Sect. 1.5.

The gauge-fixed Polyakov path-integral is well defined (for physicists' standards), therefore we take it as our starting point in constructing the theory. The original gauge symmetries reflect in the gauge-fixed action as *BRST invariance* which is a major tool in modern quantization of the string.

1.5 The Weyl Anomaly

We saw in Sect. 1.3.2 that the manifestly unitary light-cone quantization preserves Lorentz symmetry only when $d = 26$. Since the modern quantization is manifestly covariant, the pathology for $d \neq 26$ should present itself in a different way: in facts it arises from an anomaly in the Weyl symmetry.²⁸

Poincaré and **Diff** symmetries are anomaly-free, since we may regularize the theory *à la* Pauli-Villars [48] while preserving these symmetries. But there is no regulator preserving the **Weyl** symmetry, so here there is room for a quantum anomaly.

The Weyl variation of the 2d path-integral is

$$\delta \log Z[g] \Big|_{g=e^{2\phi}h} = -\frac{1}{2\pi} \int d^2\sigma \sqrt{g(\sigma)} \langle T(\sigma)^a_a \rangle \delta\phi(\sigma) \quad (1.91)$$

so Weyl invariance requires $T(\sigma)^a_a = 0$ as a *quantum operator*. While the energy-momentum tensor is classically traceless, in the presence of a Weyl anomaly the equation $T^a_a = 0$ does *not* hold quantum mechanically, and the Virasoro constraint $T_{ab} = 0$ cannot be consistently implemented (cf. footnote 16).

The trace operator T^a_a must be **Poincaré** and **Diff** invariant, since these symmetries are non-anomalous. The free theory (1.90) may be quantized in a conformal invariant way in a flat spacetime, hence T^a_a should vanish when the 2d metric is flat, up to non-universal contributions which correspond to different choices of local counter-terms. The universal part of T^a_a then depends only on the 2d Riemann curvature and its covariant derivatives. In 2d the only independent component of the Riemann tensor is the scalar curvature R . Under a change in the overall normalization of the metric $g_{ab} \rightarrow \lambda g_{ab}$, a scalar quantity $A_{r,n}$, formed by $2n$ covariant derivatives acting (in any way) on r scalar curvatures, scales as $A_{r,n} \rightarrow \lambda^{-r-n} A_{r,n}$, while $T^a_a \rightarrow \lambda^{-1} T^a_a$. We conclude that the trace of the energy-momentum tensor should have the form

²⁸ An anomaly in the Weyl symmetry implies that the covariant and light-cone quantizations are not equivalent since the classical gauge transformation which relates the two “gauges” is not a symmetry at the quantum level.

$$T^a_a = k(R + \text{const.}) \quad (1.92)$$

for some constant k which depends on the 2d theory (the second term inside the parenthesis, which survives in flat space, is the aforementioned non-universal term).

We shall not compute the constant k at this point, since it is a particular instance of a more general and interesting story which we wish to present in the proper conceptual context. Here we just quote the formula for the trace anomaly of free fields with spin λ which we shall derive and interpret in Sect. 2.5.2.

Claim 1.1 *Suppose the 2d QFT consists of a complex **free** chiral ($\equiv$ left-moving) field ψ which transforms under local holomorphic parametrizations $z' = z'(z)$ as a holomorphic λ -differential and its dual field χ transforming as a $(1 - \lambda)$ -differential*

$$\psi'(dz')^\lambda = \psi(dz)^\lambda, \quad \chi'(dz')^{1-\lambda} = \chi(dz)^{1-\lambda} \quad (1.93)$$

which are governed by the free (chiral) Dirac-like action²⁹

$$\frac{1}{2\pi} \int_{\Sigma} \psi \bar{\partial} \chi, \quad (1.94)$$

then the Weyl anomaly is

$$T^a_a = \pm \frac{1}{12} (1 - 3(1 - 2\lambda)^2)(R + \text{const.}) \quad (1.95)$$

where the upper/lower sign corresponds to ψ, χ of bosonic/fermionic statistics.

The string Faddeev–Popov ghost system b, c is a free fermionic theory of this kind with $\lambda = 2$ (together with a second right-moving copy $\tilde{b}, \tilde{c}$); see Eq. (1.86). Each matter field X^μ ($\mu = 0, \dots, d-1$) can also be decomposed in two copies (one left- and one right-moving) of *real* free chiral systems with $\lambda = 1$. Hence the matter sector (1.90) of the world-sheet theory is equivalent to $d/2$ copies of the *complex* bosonic free system with $\lambda = 1$.

Putting everything together, the Weyl anomaly (for the left-movers, the right-mover story is identical) is proportional to

$$-\frac{d}{2} (1 - 3(1 - 2\lambda)^2) \Big|_{\lambda=1} + (1 - 3(1 - 2\lambda)^2) \Big|_{\lambda=2} = d - 26 \quad (1.96)$$

and so the *critical dimension* in which the anomaly cancels and the Weyl factor $\exp[2\phi(\sigma)]$ in the metric $e^{2\phi(\sigma)} h_{ab}$ drops out from the path integral measure is $d = 26$, in agreement with the light-cone analysis in Sect. 1.3.2.

²⁹ The system (1.94) is invariant under the duality $\lambda \leftrightarrow 1 - \lambda$. Standard spin- $\frac{1}{2}$ 2d Weyl fermions correspond to the self-dual case $\lambda = 1 - \lambda = \frac{1}{2}$.

1.5.1 ★ Strings in Non-critical Dimensions

In 2d the Ricci tensor³⁰ is

$$R_{z\bar{z}} = -\partial_z \partial_{\bar{z}} \log g_{z\bar{z}}, \quad (1.97)$$

which for $g_{z\bar{z}} = e^{2\phi} \delta_{z\bar{z}}$ yields

$$\sqrt{g} R = e^{2\phi} (-2e^{-2\phi} \partial_z \partial_{\bar{z}} \phi) = -2 \partial_z \partial_{\bar{z}} \phi. \quad (1.98)$$

Returning to Eq. (1.91)

$$\begin{aligned} \delta \log Z &= -\frac{1}{2\pi} \int d^2\sigma \sqrt{g} \delta\phi(\sigma) \langle T(\sigma)^a_a \rangle = \frac{d-26}{24\pi} \int d^2\sigma \sqrt{g} \delta\phi(\sigma) (R(\sigma) + 2\mu^2) = \\ &= \frac{d-26}{12\pi} \int d^2\sigma \delta\phi(\sigma) (-\partial_z \partial_{\bar{z}} \phi + 2\mu^2 e^{2\phi}) = \frac{d-26}{12\pi} \delta \int d^2\sigma \left(\frac{1}{2} \partial_{\bar{z}} \phi \partial_z \phi + \mu^2 e^{2\phi} \right), \end{aligned} \quad (1.99)$$

that is,

$$\log Z = \frac{d-26}{12\pi} \int d^2\sigma \left(\frac{1}{2} \partial_{\bar{z}} \phi \partial_z \phi + \mu^2 e^{2\phi} \right), \quad (1.100)$$

so that the partition function for the string in non-critical dimension becomes [19]

$$\int [d\phi] \exp \left\{ -\frac{26-d}{12\pi} \int d^2\sigma \left(\frac{1}{2} \partial_{\bar{z}} \phi \partial_z \phi + \mu^2 e^{2\phi} \right) \right\} \quad (1.101)$$

which is the quantum *2d Liouville model* (see, for example, [49, 50]). We conclude that in *non-critical* dimension, $d < 26$, an extra scalar d.o.f. propagates on the world-sheet, namely the Liouville scalar field ϕ .

1.6 Ghost Zero-Modes: $\text{Aut}(\Sigma)$ and WP Moduli Geometry

It remains to discuss the crucial issue of ghost zero-modes. We refer back to **BOX 1.4** for basic properties of 2d manifolds. We have two kinds of zero-modes:

- zero-modes of the c ghost
- zero-modes of the b ghost.

They have different geometric origin.

Zero-Modes of the c Ghost

The Faddeev–Popov ghost $c \equiv c^z$ is a vector field (the parameter of an infinitesimal diffeomorphism) with the wrong (fermionic) statistics. Its equations of motion are

$$\bar{\partial}c = 0 \quad \begin{array}{l} \text{equivalent to } \mathfrak{L}_c g = 2\lambda g \\ \text{for some function } \lambda, \end{array} \quad (1.102)$$

³⁰ This is just the formula for the Ricci curvature in a Kähler manifold [38, 44] to be reviewed in Sect. 11.1.2; in 2d all metrics are (locally) Kähler; see **BOX 1.4**.

so, when Σ has no boundary, a zero-mode of c is nothing else than a *global* conformal motion, i.e. an infinitesimal holomorphic automorphism of Σ , that is, an element of the Lie algebra $\text{aut}(\Sigma)$ of the *complex* Lie group $\text{Aut}(\Sigma)$; see Table 1.1. The $\tilde{c}$ zero-modes are the elements of the complex conjugate Lie algebra $\text{aut}(\Sigma)^*$.

Consider now a world-sheet with boundary, $\partial\Sigma \neq \emptyset$. In Exercise 1.2 the reader was asked to determine the appropriate boundary conditions for the b, c ghosts. The short answer is that the correct b.c. are such that the zero-modes of c describe global (infinitesimal) automorphisms of Σ which preserve set-wise the boundary $\partial\Sigma$. More specifically, when $\partial\Sigma \neq \emptyset$ the boundary condition couples c and $\tilde{c}$ which are no longer independent fields (see Sect. 2.6 for details). This coupling corresponds to the geometric fact that now $\text{Aut}(\Sigma)$ is a *real* Lie group³¹ (cf. Table 1.2) and the two Lie algebras $\text{aut}(\Sigma)$ and $\text{aut}(\Sigma)^*$ are identified. The total number of $c, \tilde{c}$ zero-modes is then $\dim_{\mathbb{R}} \text{Aut}(\Sigma)$, and the $c, \tilde{c}$ zero-modes are independent (resp. identified) if the Lie group $\text{Aut}(\Sigma)$ is complex (resp. real).

For instance: on the sphere we have 3 zero-modes for c and 3 zero-modes for $\tilde{c}$, while on the torus there are one zero-mode for c and one for $\tilde{c}$. In the disk we have only 3 *real* zero-modes which involve both ghost fields.

We describe the situation in the field-theoretic language. The conformal gauge fixing (1.80) does not eliminate all gauge freedom: there remains a *finite-dimensional Lie group*³² of residual gauge symmetries, namely the holomorphic automorphism group $\text{Aut}(\Sigma)$. To fully implement the Faddeev–Popov procedure, we need to fix also the residual $\text{Aut}(\Sigma)$ gauge. The gauge fixing of the finite-dimensional residual gauge symmetry then introduces a *finite-dimensional* Faddeev–Popov determinant.

A convenient way to proceed is as follows. Let us consider first the case of Σ a closed Riemann surface. Let $m = \dim_{\mathbb{C}} \text{Aut}(\Sigma)$ be the number of c zero-modes ($\equiv$ the number of $\tilde{c}$ ones). Choose m distinct points

$$z_k \in \Sigma \quad (k = 1, \dots, m) \quad (1.103)$$

and fix their coordinates to some reference value $\hat{z}_k$. Each condition $z_k = \hat{z}_k$ gives one complex (or 2 real) conditions: by counting parameters we see that fixing m points kills all continuous residual symmetries, that is, the gauge symmetries which still remain unfixed, if any, form, at most, a *finite* group.

Explicitly: we take 3 distinct points on the sphere $\mathbb{P}^1$ and fix them to be at 0, 1, ∞ , or, in the torus case ($g = 1$), we fix one point to be at the origin. For $g > 1$ the c field has no zero-modes.

³¹ As it will be more clear in Chap. 2, this is a consequence of the *Schwarz principle* in complex analysis [51]. The point is that given a complex surface Σ with boundary we may construct a *closed double* Σ^c such that $\Sigma = \Sigma^c / C$ where $C: \Sigma^c \rightarrow \Sigma^c$ is an involutive *anti*-holomorphic automorphism of Σ^c such that $\partial\Sigma$ is identified with the fixed set of C . On the closed double Σ^c we have independent $c, \tilde{c}$ zero-modes, while (by construction) the space of zero-modes on Σ is the C -invariant subspace of the zero-modes on Σ^c , $\text{aut}(\Sigma) = \text{aut}(\Sigma^c)^C$, which is obviously defined over $\mathbb{R}$. Now $C: c \leftrightarrow \tilde{c}$ (since C maps left-movers to right-movers), and hence the C -invariant subspace does not split into independent zero-mode spaces for c and $\tilde{c}$.

³² This is an actual Lie group [46].

Exercise 1.3 Show that the only automorphism of the sphere fixing 0, 1, and ∞ is the identity. **Hint:** An automorphism of the Riemann sphere $f: \mathbb{P}^1 \rightarrow \mathbb{P}^1$ which fixes ∞ is a degree 1 polynomial $f(z) = az + b$.

The automorphism of the torus³³ fixing the origin forms a finite group; for generic tori this is just the $\mathbb{Z}_2$ group $z \rightarrow \pm z$. Of course, to avoid overcounting, one has to divide the path integral by the order of the *generic* finite group.

Complete Gauge Fixing

The full set of gauge conditions now is

$$g(z)\bar{z}^z = 0 \quad \text{and} \quad z_k - \hat{z}_k = 0 \quad k = 1, \dots, m, \quad \text{plus complex conjugates} \quad (1.104)$$

whose infinitesimal **Diff** gauge variations of parameter $c(z)$ are

$$\bar{\partial}c(z) \quad \text{and} \quad c(z_k) \quad k = 1, \dots, m. \quad (1.105)$$

Introducing dual anti-ghosts for all gauge-fixing conditions (1.104), $b(z)$, and, respectively, η_k ,³⁴ we get *formally* (that is, assuming—for the moment—that the anti-ghost $b(z)$ has no zero-modes) the Faddeed–Popov determinant in the form

$$\det P = \Delta_{\text{FP}} \Delta_{\text{FP}}^* \quad (1.106)$$

where the factor Δ_{FP} (resp. Δ_{FP}^*) is the path integral over the left-moving (resp. right-moving) FP ghosts

$$\begin{aligned} \Delta_{\text{FP}} &= \int [db dc] \prod_{k=1}^m d\eta_k \exp \left(- \sum_k \eta_k c(z_k) - \frac{1}{2\pi} \int b(z) \bar{\partial}c(z) \right) = \\ &= \int [db dc] c(z_1) c(z_2) \dots c(z_m) \exp \left(- \frac{1}{2\pi} \int b(z) \bar{\partial}c(z) \right). \end{aligned} \quad (1.107)$$

The result we got has a simple interpretation: in the presence of m c -ghost zero-modes one needs to insert m Fermi fields $c(z)$ in the path integral in order to *soak up* the zero-modes and get a non-zero answer. The choice of soaking up the zero-mode by inserting the ghost field at m distinct points is a very convenient one since it does not lead to extra finite Jacobians in the functional measure, except for the factor automatically produced by the m -point function of the field $c(z)$, as Eq. (1.107) shows.

Open String In the open case, the situation is similar, except that $\text{Aut}(\Sigma)$ is a real Lie group instead of a complex one, and to fix the residual symmetry up to finite groups we need $m \equiv \dim_{\mathbb{R}} \text{aut}(\Sigma)$ *real* gauge conditions. The most convenient procedure is to fix m point *on the boundary*: since the boundary has real dimension 1, this gives

³³ We see a torus as the quotient $\mathbb{C}/\Lambda$ where $\Lambda \subset \mathbb{C}$ is a lattice. z is the complex coordinate of $\mathbb{C}$.

³⁴ $b(z)$ is a local chiral field dual to $\bar{\partial}c(z)$, while the η_k are global Grassmann parameters.

the right number of gauge-fixing conditions. The rule is to insert $c(x_i)$ at m distinct points in the boundary $\partial\Sigma$, where m is the real dimension of $\text{Aut}(\Sigma)$, that is, 3 for the disk, 1 for the annulus, etc.

1.6.1 The Riemann–Roch Theorem

The story above is reminiscent of instantons in 4d QCD [53–55]. Fermions have zero-modes in a gauge background with non-zero instanton number. To get non-zero amplitudes, we need to soak up the zero-modes by inserting fermions in the path integral and the result is a chiral condensate of quark fields. The number of zero-modes in an instanton background is given by the *Atiyah–Singer index theorem* [31, 56, 57] for the Dirac operator $\mathcal{D}$ coupled to the Yang–Mills field. The index theorem may be understood in terms of the 4d Adler–Bardeen axial anomaly [58]

$$\partial^\mu J_{\mu 5} \propto \text{tr}(F \tilde{F}). \quad (1.108)$$

Integrating this equation over spacetime, we get the net variation of chirality equal to the difference $n_+ - n_-$ in the numbers n_+, n_- of zero-modes with positive and negative chirality. The variation is proportional to the instanton topological charge $\nu \equiv \int \text{tr}(F \tilde{F})/(32\pi^2)$

$$n_+ - n_- = 2N_f \nu \quad \text{in QCD with } N_f \text{ flavors}, \quad (1.109)$$

which is the Atiyah–Singer theorem for $\mathcal{D}$. In two dimensions the situation is similar, but much easier. The action for the ghosts is a kind of first-order 2d Dirac action

$$\int b \bar{\partial} c \quad (1.110)$$

which differs from the standard 2d chiral Dirac action only because b, c transform in a different way under **Diff**, that is, because they are sections of line bundles distinct from the spin bundles where the usual spin- $\frac{1}{2}$ Weyl fermions take value. As in QCD, we have a chiral $U(1)$ current $bc(z)$ (whose charge is the *ghost number*) which is classically conserved, but not quantum mechanically. The $U(1)$ anomaly implies an index theorem just as in Eq. (1.109), and the index theorem will give us a formula for the number of b, c zero-modes:

$$\begin{aligned} \text{violation of ghost number} &= \#(c\text{-zero-modes}) - \#(b\text{-zero-modes}) = \\ &= \text{topological invariant.} \end{aligned} \quad (1.111)$$

The 2d index theorem is known as the **Riemann–Roch theorem** [38, 59, 60]. The theorem can be proven by the usual techniques of Algebraic Geometry, or by one-

BOX 1.6 - Basic facts about line bundles on Riemann surfaces

A Hermitian (complex) line bundle $\mathcal{L} \rightarrow \Sigma$ is a vector bundle with fiber $\simeq \mathbb{C}$ endowed with a Hermitian norm along the fibers. Let $\cup_\alpha U_\alpha$ be a sufficiently fine open cover of Σ . In a local trivialization $\mathcal{L}|_{U_\alpha} \simeq U_\alpha \times \mathbb{C}$ we write ψ_α for the complex coordinate along the fiber; the squared-norm has the form $\|\psi_\alpha\|^2 = h_\alpha |\psi_\alpha|^2$ for a smooth positive function h_α . On the overlap $U_\alpha \cap U_\beta$, $\psi_\alpha = \lambda_{\alpha\beta} \psi_\beta$ for $\lambda_{\alpha\beta}$ a complex function nowhere vanishing in $U_\alpha \cap U_\beta$. $\lambda_{\alpha\beta}$ satisfies the *cocycle identities* [38, 43]

$$\lambda_{\alpha\beta} \lambda_{\beta\alpha} = 1, \quad \lambda_{\alpha\beta} \lambda_{\beta\gamma} \lambda_{\gamma\alpha} = 1. \quad *$$

Conversely a $\lambda_{\alpha\beta}$ satisfying (*) defines a line bundle. On $U_\alpha \cap U_\beta$ we have $h_\beta = h_\alpha |\lambda_{\alpha\beta}|^2$, so that $\|\psi_\alpha\|_\alpha^2 = \|\psi_\beta\|_\beta^2$ and the norm is independent of the trivialization. The cocycles, hence the line bundles, form an Abelian group under multiplication. Two line bundles are isomorphic iff $\psi'_\alpha = \lambda_\alpha \psi_\alpha$ for λ_α a function nowhere vanishing in U_α . A cocycle is called a coboundary if has the form $\lambda_{\alpha\beta} = \lambda_\alpha \lambda_\beta^{-1}$. The isomorphism classes of line bundles then form a group isomorphic to the group of cocycles modulo the group of coboundaries. Let $\nabla = d + A$ be a (Abelian) gauge connection on $\mathcal{L}$ which is a metric for the Hermitian structure, i.e. $d\langle\psi, \eta\rangle = \langle\nabla\psi, \eta\rangle + \langle\psi, \nabla\eta\rangle$. The 2-form $(d + A)^2$ is the curvature of A ; in a complex manifold of dimension 1 all 2-forms are of pure type (1, 1), so $(d + A)^2|_{(0,2)} = 0$. Let $\nabla = D + \bar{D}$ be the decomposition of ∇ into (1, 0) and (0, 1) parts: $\bar{D}^2$ is the (0, 2) part of the curvature, hence zero. Then

Proposition 1.1 *All smooth Hermitian line bundles on a Riemann surface are holomorphic, i.e. there is a trivialization such that the metric connection takes the form (Chern connection)*

$$D = h^{-1} \partial h, \quad \bar{D} = \bar{\partial}.$$

In the holomorphic trivialization, the cocycle $\lambda_{\alpha\beta}$ is a nowhere vanishing holomorphic function. The group of isomorphism classes of holomorphic line bundles, called the *Picard group* $\text{Pic}(\Sigma)$, is isomorphic to the group of such holomorphic cocycles modulo the coboundaries, that is to $H^1(\Sigma, \mathcal{O}^\times)$ [43]. The neutral element in the group is the trivial line bundle $\mathcal{O}$ of holomorphic functions. The exact sequence of sheaves [43]

$$0 \rightarrow \mathbb{Z} \rightarrow \mathcal{O} \xrightarrow{\exp(2\pi i \cdot)} \mathcal{O}^\times \rightarrow 1$$

yields the Picard group as an extension of well-known Abelian groups [43]

$$0 \rightarrow H^1(\Sigma, \mathcal{O})/H^1(\Sigma, \mathbb{Z}) \rightarrow H^1(\Sigma, \mathcal{O}^\times) \xrightarrow{c} H^2(\Sigma, \mathbb{Z}) \cap H^{1,1}(\Sigma, \mathbb{R}) \rightarrow 0 \quad *.$$

The map c is the Chern class; explicitly it is given by the curvature (1,1) divided by 2π

$$c(\mathcal{L}) = \frac{1}{2\pi i} \partial \bar{\partial} \log h.$$

The degree of a line bundle is

$$\deg(\mathcal{L}) \equiv \int_\Sigma c(\mathcal{L}) \in \mathbb{Z}.$$

The degree (and the Chern class) are group homomorphisms $\text{Pic}(\Sigma) \rightarrow \mathbb{Z}$

$$\deg \mathcal{L}^{-1} = -\deg \mathcal{L}, \quad \deg(\mathcal{L} \otimes \mathcal{L}') = \deg \mathcal{L} + \deg \mathcal{L}'.$$

The group of holomorphic line bundles of degree 0, $\text{Pic}(\Sigma)^0$, is the *Jacobian* $J(\Sigma)$ of Σ

$$J(\Sigma) \equiv \ker c = H^1(\Sigma, \mathcal{O})/H^1(\Sigma, \mathbb{Z}) \quad \text{from } (*)$$

which is a complex torus (in fact an Abelian variety) of complex dimension $\dim_{\mathbb{C}} H^1(\Sigma, \mathcal{O}) = g$ (this is the *definition* of the genus g of Σ).

Canonical Bundle K An example of Hermitian line bundle is the holomorphic tangent bundle $T\Sigma$ whose sections are vector fields of type $(1, 0)$, $f^z \partial_z$. By definition, its Chern class is represented by the Ricci form divided by 2π and its degree is the Euler characteristic $\chi = 2 - 2g$. The inverse (dual) bundle $T\Sigma^{-1}$ is called the *canonical line bundle* K of Σ . It is the bundle of $(1, 0)$ differential forms, $f_z dz$. Its degree is $\deg K = -\deg T\Sigma = 2g - 2$. K is the *dualizing bundle* (or rather *sheaf* [52]). This means that, for any line bundle $\mathcal{L}$,

$$H^1(\Sigma, \mathcal{L}) \cong H^0(\Sigma, K \otimes \mathcal{L}^{-1})^\vee \quad \text{Serre duality theorem [38, 46, 61].}$$

loop perturbation theory via the Adler–Bardeen theorem (see below). We shall give the third proof in the next chapter.

For rudiments about line bundles on Riemann surfaces, see **BOX 1.6**; for the translation of these facts from the framework of Analytic/Algebraic Geometry to the usual language of QFT in terms of gauge fields, field strengths, fluxes, and wave-functions, see **BOX 1.7**. As explained in **BOX 1.7**, the following two Gaussian quantum field theories are equivalent:

- a free 2d theory with a Dirac-like quadratic Lagrangian $b\bar{\partial}c$, where c (resp. b) is a λ -differential (resp. a $(1 - \lambda)$ -differential), i.e. a section³⁵ of the line bundle K^λ (resp. $K^{1-\lambda}$) of degree $\deg K^\lambda = \lambda \deg K \equiv \lambda(2g - 2)$;
- an *ordinary* spin- $\frac{1}{2}$ Weyl fermion ψ , i.e. a section³⁶ of $K^{1/2}$ *coupled to* a background gauge field $\bar{A}$ which is a $U(1)$ connection on $K^{\lambda-1/2}$ with Lagrangian

$$\bar{\psi}(\bar{\partial} + \bar{A})\psi, \tag{1.112}$$

where the gauge field $\bar{A}$ is $(\frac{1}{2} - \lambda)$ times the Levi-Civita connection $\bar{A}_{\text{tan}}$ on the tangent bundle (written in a unitary frame; see **BOX 1.7**).

By the usual Adler–Bardeen formula for ordinary 2d fermions in a gauge background, the $U(1)$ anomaly of the second (hence of the first) theory is

³⁵ Genuine global line bundles exist only for λ integer or half-integer; at the formal level we can work with λ real. See **BOX 1.10**.

³⁶ The precise meaning of taking a square-root $K^{1/2}$ of K is explained in **BOX 1.10**.

$$\begin{aligned} \#(\text{zero modes of } \psi) - \#(\text{zero modes of } \bar{\psi}) &= \\ &= \left(\frac{1}{2} - \lambda\right) \frac{1}{2\pi} \int_{\Sigma} d\bar{A}_{\text{tan}} = \left(\frac{1}{2} - \lambda\right)(2 - 2g) = 1 - g + \deg(K^{\lambda}). \end{aligned} \quad (1.113)$$

Since the zero-modes of c are the global holomorphic sections of $\mathcal{L} \equiv K^{\lambda}$ over Σ , whose vector space is written in geometry as $H^0(\Sigma, \mathcal{L})$ [38, 45], while the zero-modes of b are the global holomorphic sections of $K\mathcal{L}^{-1} \equiv K^{1-\lambda}$, whose space is $H^0(\Sigma, K\mathcal{L}^{-1})$, we can rewrite the result in the standard geometric form as follows.

Theorem 1.5 (Riemann–Roch [38, 59, 60]) *Σ a compact Riemann surface of genus g (a smooth projective curve over $\mathbb{C}$ of genus g), $\mathcal{L}$ a holomorphic line bundle. Then*

$$\dim H^0(\Sigma, \mathcal{L}) - \dim H^0(\Sigma, K\mathcal{L}^{-1}) = 1 - g + \deg \mathcal{L}. \quad (1.114)$$

BOX 1.7 - Hermitian line bundles as $U(1)$ gauge fields

Hermitian line bundles may be seen in a “more physical” language as Abelian $U(1)$ gauge fields on a Riemann surface. As we saw in BOX 1.6, the norm-squared of a section ψ has (locally) the form $h|\psi|^2$ for some positive function h . We can perform a complex $GL(1, \mathbb{C})$ gauge transformation $\psi \rightarrow h^{1/2}\psi$ to a unitary trivialization where the norm is simply $|\psi|^2$. A metric connection in the unitary trivialization has the form $d + \mathcal{A}$ with $\mathcal{A}$ a one-form with value in the Lie algebra $\mathfrak{u}(1)$, i.e. an (anti-Hermitian) Abelian connection. The complex gauge transformation of the Chern connection in BOX 1.6 gives the equivalent unitary gauge connection [43, 45]

$$d + \mathcal{A} = d + \frac{1}{2} \partial \log h - \frac{1}{2} \bar{\partial} \log h.$$

The curvature is the same as before (of course) and we see that the degree is just $1/2\pi$ times the magnetic flux through the surface Σ , i.e. the number of Dirac units of quantized flux. If the Abelian gauge field $\mathcal{A}$ is flat, i.e. has zero field strength, its gauge-invariant content is given by the monodromy representation $W: \pi_1(\Sigma) \rightarrow U(1)$ given by the Wilson lines along non-trivial loops in the surface Σ . Since $U(1)$ is Abelian, the monodromy representation factors through the Abelianization $\pi_1(\Sigma)^{\text{Ab}}$ of the fundamental group

$$\pi_1(\Sigma)^{\text{Ab}} \stackrel{\text{def}}{=} \pi_1(\Sigma) / [\pi_1(\Sigma), \pi_1(\Sigma)] \equiv H_1(\Sigma, \mathbb{Z}) \simeq H^1(\Sigma, \mathbb{Z})^{\vee},$$

so that the gauge-inequivalent configurations of an Abelian gauge field with zero field strength on a compact surface Σ is given by

$$\text{Hom}(H^1(\Sigma, \mathbb{Z})^{\vee}, U(1)) \equiv J(\Sigma) \equiv \text{the Jacobian.}$$

Note 1.6 The Riemann–Roch theorem counts the number of quantum states of a charged electron moving on a surface with a magnetic flux $2\pi \deg \mathcal{L}$ which belong to the lowest Landau level, so it is the same thing as the Heisenberg indetermination principle (seeing Σ as a phase space).

Application to b, c ghosts The string ghosts c, b correspond to $\lambda = -1$ so that

$$\#(c \text{ zero-modes}) - \#(b \text{ zero-modes}) = 1 - g + 2 - 2g = 3 - 3g. \quad (1.115)$$

That is,

- for $g = 0$ (sphere), we have 3 c zero-modes and no b zero-modes;
- for $g = 1$ (torus), we have 1 zero-mode for c and 1 for b (the constant modes);
- for $g > 1$, there are no c zero-modes and $3g - 3$ b zero-modes.

1.6.2 b Zero-Modes and the Moduli Space

Zero-modes of b are the traceless symmetric tensors b_{ab} such that (cf. Eq. (1.82))

$$D^a b_{ab} \equiv (P^\dagger b)_b = 0, \quad (1.116)$$

i.e. symmetric 2-tensors which are orthogonal to all **Diff** + **Weyl** infinitesimal deformations of the metric. In other words: ***the b zero-modes are infinitesimal deformations of the world-sheet metric which cannot be obtained by gauge transformations.*** Put differently, a b zero-mode corresponds to a deformed metric

$$ds^2 = g_{z\bar{z}} dz \otimes d\bar{z} + \epsilon b_{z\bar{z}} dz \otimes dz + \bar{\epsilon} \bar{b}_{\bar{z}\bar{z}} d\bar{z} \otimes d\bar{z} \quad (1.117)$$

which is ***note*** gauge-equivalent to the original one $ds^2 = g_{z\bar{z}} dz \otimes d\bar{z}$.

Let us take a different viewpoint. We have seen in **BOX 1.4** that a conformal structure is the same as a complex structure. All surfaces of given genus g are *diffeomorphic* (i.e. equivalent in the C^∞ sense), but the unique underlying smooth surface may admit several *inequivalent* complex structures. The complex structures of surfaces of genus g are parametrized by a *finite-dimensional* complex manifold $\mathcal{M}_g$, called their *complex moduli space* [61, 62]. The above discussion shows

$$\dim_{\mathbb{C}} \mathcal{M}_g \equiv \#(b \text{ zero-modes}) = \begin{cases} 0 & g = 0 \\ 1 & g = 1 \\ 3g - 3 & g \geq 2. \end{cases} \quad (1.118)$$

More generally, we may consider the moduli space $\mathcal{M}_{g,n}$ of surfaces of genus g with n *punctures*, that is, with n distinct marked points. Having eliminated from the path integral, the gauge redundancies associated with **Diff** + **Weyl** by fixing the gauge, we are left with an integral over the finite-dimensional space $\mathcal{M}_{g,n}$ of gauge *inequivalent* geometries

$$\dim_{\mathbb{C}} \mathcal{M}_{g,n} = \max\{3g - 3 + n, 0\} \quad \text{for } n \geq 1. \quad (1.119)$$

In the language of instanton physics, the moduli are the “collective coordinates” we have to integrate over, and—as in the instanton calculus—the finite-dimensional

Faddeev–Popov determinant associated with the b zero-modes is the Jacobian which produces the correct measure on the moduli space $\mathcal{M}_{g,n}$.

From Eq. (1.117) we see that the space of b zero-modes $H^0(\Sigma, K^2) \cong H^1(\Sigma, K^{-1})^*$ is the holomorphic³⁷ tangent space to $\mathcal{M}_g$

$$H^0(\Sigma_g, K^2) \cong T_{\Sigma_g} \mathcal{M}_g. \quad (1.120)$$

The zero-modes of b , i.e. the *holomorphic* $b_{z\bar{z}}$ such that $\bar{\partial}_{\bar{z}} b_{z\bar{z}} = 0$ are called *Beltrami differentials* (a.k.a. *quadratic differentials*). The mathematically minded reader will find in **BOX 1.8** a more detailed explanation of why the Beltrami differentials correspond to infinitesimal deformations of the complex structure of Σ .

In **BOX 1.9** we show that for each complex structure on a Riemann surface of genus $g \geq 2$, there is a *unique*³⁸ metric satisfying

$$R_{z\bar{z}} = -g_{z\bar{z}}. \quad (1.121)$$

Hence the moduli space $\mathcal{M}_g$ may be identified with the space of normalized Einstein metrics (up to isometry). In other words: **in each Diff + Weyl gauge equivalence class of Riemannian metrics on a compact 2-manifold there is precisely one Einstein metric normalized to volume 1**. We fix our reference metric h_{ab} in Eq. (1.80) to be (a convenient multiple of) this special metric.

Moduli Space and its Volume Form

As stated above, the finite “Faddeev–Popov” determinant associated with the $b, \tilde{b}$ zero-modes is the Jacobian between the naive functional measure and the correct measure in the complex moduli space. We write m^i for local coordinates in $\mathcal{M}_g$ [61, 62]. The resulting measure on the moduli is the volume form

$$\frac{\Omega^{\dim \mathcal{M}_g}}{\dim \mathcal{M}_g!} \equiv \frac{1}{\dim \mathcal{M}_g!} \left(i G_{j\bar{k}} dm^j \wedge d\bar{m}^k \right)^{\dim \mathcal{M}_g} \quad (1.122)$$

induced by the Hermitian (in facts Kähler) metric on $\mathcal{M}_g$

$$ds^2 = G_{j\bar{k}} dm^j d\bar{m}^k \equiv \left(\int_{\Sigma} \sqrt{g} d^2\sigma g^{\alpha\gamma} g^{\beta\delta} \partial_{m^j} g_{\alpha\beta} \partial_{\bar{m}^k} g_{\gamma\delta} \right) dm^j d\bar{m}^k \quad (1.123)$$

which (up to conventional factors of 2) is just the *Weil–Petersson metric* on $\mathcal{M}_g$; see **BOX 1.8**. The same measure may be written in terms of insertions of the fields $b, \tilde{b}$ in the path integral to “soak up” their zero-modes

³⁷ **Warning** The complex structure on $\mathcal{M}_g$ used in string theory is the *opposite* of the conventional one in Kodaira–Spencer theory, **BOX 1.8**, where $(T_{\Sigma_g} \mathcal{M}_g)_{\text{KS}} \cong H^1(\Sigma_g, K^{-1}) \cong H^0(\Sigma_g, K^2)^*$.

³⁸ This may be seen as a baby instance of the Yau theorem [71] about existence and uniqueness of Einstein–Kähler metrics in compact complex manifolds.

BOX 1.8 - Rudiments of Kodaira–Spencer theory [63]

Suppose we have a compact smooth manifold X of even dimension and assume it admits a reference complex structure—we write X_0 for the complex manifold obtained by endowing X with this complex structure. We wish to construct all other *inequivalent* complex structures which X can have. We end up with a family X_t of complex manifolds parametrized by $t \in \mathcal{M}$, where $\mathcal{M}$ is the *moduli space*. Equivalently we get a proper fibration $X \rightarrow \mathcal{M}$ whose fibers are the X_t . We consider continuous deformations of the complex structures, so $\mathcal{M}$ is connected. All fibers in X are diffeomorphic; then the complex coordinates z^i of X_0 are C^∞ coordinates for all X_t ’s. Specifying a complex structure on a smooth manifold X is equivalent to specifying which *local* C^∞ complex functions are holomorphic (technically: specifying the sheaf $\mathcal{O}$ of germs of holomorphic functions as a subsheaf of the sheaf of germs of C^∞ complex functions $\mathcal{A}$). A smooth function f is holomorphic iff it satisfies the Cauchy–Riemann equation $\bar{\partial} f = 0$, so deforming the complex structure is the same as deforming the Dolbeault operator $\bar{\partial}$ which is a first-order differential operator from the functions to the $(0, 1)$ -forms vanishing on the constants. The local holomorphic functions on X_t are the kernel of the deformed $\bar{\partial}_t$, i.e. $f \in \mathcal{O}_t \Leftrightarrow \bar{\partial}_t f = 0$. Since we are only interested in the kernel of $\bar{\partial}_t$, we may reduce to operators of the form

$$\bar{\partial}_t = d\bar{z}^i (\partial_{\bar{z}^i} - \phi(t)^j_{\bar{z}^i} \partial_{z^j})$$

where $\phi(t) \equiv d\bar{z}^i \phi(t)^j_{\bar{z}^i} \partial_{z^j}$ is the *Kodaira–Spencer vector*. It depends on the moduli t^a holomorphically. We need the kernel to contain enough local holomorphic functions to form a holomorphic local coordinate systems. This entails an integrability condition

$$\bar{\partial}_t^2 = 0 \quad \Leftrightarrow \quad \bar{\partial}\phi(t) + \frac{1}{2}[\phi(t), \phi(t)] = 0$$

called the *Kodaira–Spencer (KS) equation*. It has the form of a “zero field strength” equation. There are trivial solutions, i.e. “pure gauge” $\phi = e^{-\xi} \bar{\partial} e^\xi$ where ξ is a smooth $(1, 0)$ vector field. The trivial solutions do not correspond to deformations of the complex structure, just to writing the same complex structure in different coordinates. The (finite) deformations are given by the solutions to the KS equation modulo the trivial ones. We consider now the infinitesimal deformations, i.e. $\phi(\epsilon) = \epsilon\phi_1 + \epsilon^2\phi_2 + \dots$ and $\xi = \epsilon\xi_1 + \dots$. To the leading order we find

$$\bar{\partial}\phi_1 = 0, \quad \phi_1 \simeq \phi_1 + \bar{\partial}\xi_1,$$

so formally the infinitesimal deformations are given by $H^1(X, TX)$. However in general, not all formal infinitesimal deformations can be extended to actual deformations since the KS equation may be *obstructed* in higher order. This cannot happen in one complex dimension (i.e. for Riemann surfaces): the KS equation is trivially satisfied because its LHS is a $(0, 2)$ -form (with $(1, 0)$ vector coefficients) hence identically zero in dimension 1. We remain with the condition that ϕ is not “pure gauge”; by standard Hodge theory each cohomology class has a unique harmonic representative, so the holomorphic tangent space to the moduli $\mathcal{M}$ is given by the harmonic KS vectors, i.e. the Beltrami differentials, $\partial_{\bar{z}}(\phi_{\bar{z}} g_{z\bar{z}}) = 0$, which is the $\bar{b}$ zero-mode equation. The complex dimension of the moduli space $\mathcal{M}_g$ of genus- g Riemann surfaces is equal to the number of zero-modes of $\bar{b}$, i.e. 0 for $g = 0$, 1 for $g = 1$, and $3g - 3$ for $g \geq 2$.

Weil–Petersson Metric We have the inclusion map (*isomorphism* in dimension 1)

$$\iota: T_t \mathcal{M} \xrightarrow{\sim} H^1(X_t, TX_t), \quad \iota(\partial_{t^i}) = \phi_i^a \partial_{z^a}$$

which yields a natural metric on $\mathcal{M}$, the *Weil–Petersson (WP) metric*,

$$ds^2 = G_{j\bar{k}} dt^j d\bar{t}^k = dt^j d\bar{t}^k \int_{X_t} g_{a\bar{b}} \phi_j^a \wedge * \bar{\phi}_k^{\bar{b}}.$$

Fact 1. The WP metric on the moduli of curves is Kähler with non-positive Riemannian curvature operators, negative sectional curvatures, and negative holomorphic bisectional curvatures [64].

Fact 2. Let Ω be the WP Kähler form. The moduli volume is finite: its value is computed by 2d quantum gravity, and it is essentially known; see [65–70] and references therein.

BOX 1.9 - Uniformization of Riemann surfaces [59]

Theorem 1.6 (Riemann) *Up to biholomorphic equivalence, there are only three simply connected one-dimensional complex manifolds, the sphere ($\equiv$ the projective line) $\mathbb{P}^1$, the complex plane $\mathbb{C}$, and the upper half plane $\mathcal{H} \equiv \{z \in \mathbb{C} : \text{Im } z > 0\}$. In particular, all simply connected domains in $\mathbb{C}$, different from $\mathbb{C}$, are biholomorphic to the upper half-plane.*

Fact *The holomorphic automorphism group of the above three simply connected one-dimensional complex manifolds acts transitively*

$$\text{Aut}(\mathbb{P}^1) = \text{PSL}(2, \mathbb{C}), \quad \text{Aut}(\mathbb{C}) = \mathbb{C}^\times \ltimes \mathbb{C}, \quad \text{Aut}(\mathcal{H}) = \text{PSL}(2, \mathbb{R}),$$

and there is a unique (up to normalization) Kähler metric which is invariant under the full automorphism group. Equipped with this metric, the surface is a Hermitian symmetric space [72] (i.e. a symmetric Riemannian manifold whose symmetric metric is Kähler). It is convenient to normalize the metric so that $R_{z\bar{z}} = \lambda g_{z\bar{z}}$ with, respectively, $\lambda = +1, 0, -1$:

$$\frac{dz d\bar{z}}{(1 + |z|)^2}, \quad dz d\bar{z}, \quad \frac{dz d\bar{z}}{2(\text{Im } z)^2}.$$

Let Σ be a compact Riemann surface of genus g . It has the form $\Sigma = \tilde{\Sigma}/\Gamma$ where $\tilde{\Sigma}$ is the universal cover of Σ and $\Gamma \subset \text{Aut}(\tilde{\Sigma})$ is a torsionless discrete subgroup (called the *surface group* of Σ), acting freely and properly discontinuously, such that

$$\pi_1(\Sigma) \simeq \Gamma.$$

Since $\tilde{\Sigma}$ is simply connected, one must have $\tilde{\Sigma} = \mathbb{P}^1, \mathbb{C}$, or $\mathcal{H}$. By the Gauss–Bonnet theorem

$$\chi(\Sigma) \equiv 2 - 2g = \frac{1}{4\pi} \int_{\Sigma} \sqrt{g} d^2\sigma R$$

is positive, zero, or negative for $\tilde{\Sigma} = \mathbb{P}^1, \mathbb{C}$, or respectively $\mathcal{H}$. Hence

$$\tilde{\Sigma} = \begin{cases} \mathbb{P}^1 & g = 0 \\ \mathbb{C} & g = 1 \\ \mathcal{H} & g \geq 2. \end{cases}$$

Corollary 1.1 *All surfaces with $g \geq 2$ admit a unique Kähler metric such that $R_{z\bar{z}} = -g_{z\bar{z}}$.*

$$\left(\frac{1}{4\pi} \int_{\Sigma} b_{zz}(\phi_k)_{\bar{z}}^z dm^k \right)^{\dim \mathcal{M}_g} \left(\frac{1}{4\pi} \int_{\Sigma} \tilde{b}_{\bar{z}\bar{z}}(\bar{\phi}_k)_z^{\bar{z}} d\bar{m}^k \right)^{\dim \mathcal{M}_g} \quad (1.124)$$

That is, we must insert in the path integral the b -ghosts “folded” in the corresponding Beltrami differentials ϕ_k . The 2d path integral with the appropriate ghost insertions produces a differential form $\mathcal{I}_g$ in $\mathcal{M}_g$ of type $(\dim \mathcal{M}_g, \dim \mathcal{M}_g)$ with the structure

$$\mathcal{I}_g = Z_g \cdot \Omega^{\dim \mathcal{M}_g}, \quad (1.125)$$

where the function $Z_g : \mathcal{M}_g \rightarrow \mathbb{C}$ is produced by the path integral over non-zero-modes. The differential form $\mathcal{I}_g$ has the appropriate degree and type to be integrated over $\mathcal{M}_g$ to produce a number which is our quantum amplitude

$$g\text{-loop vacuum amplitude} = \int_{\mathcal{M}_g} \mathcal{I}_g. \quad (1.126)$$

The n -point amplitudes are given by a similar integral over the moduli space $\mathcal{M}_{g,n}$.

The integrand Z_g may diverge only because of infra-red singularities. As stated in **BOX 1.8**, the Weil–Petersson volume of $\mathcal{M}_{g,n}$ is *finite* [65–70]. Therefore, if the function Z_g is *bounded* in $\mathcal{M}_g$, the g -loop contribution is automatically *finite*. The corresponding quantity in QFT (the g -loop vacuum amplitude) is *UV divergent* and should be regularized and renormalized. Instead in string theory the higher loop corrections are automatically *UV finite*. In the bosonic string the answer is still divergent, but the divergence arises from the IR not the UV: we should expect an IR divergence since the theory has a tachyon and we are expanding around an unstable vacuum. This IR divergence is as an “artifact” of our poor treatment of the theory. In the superstring case—where there are no tachyons—even the IR diverges will cancel, and the quantum amplitudes will be finite to all loop orders. We shall return to these issues when explicitly computing the amplitudes in Chaps. 4, 5, and 10.

1.7 The Superstring

In the bosonic string *à la* Polyakov the basic principles are world-sheet **Diff** and **Weyl** gauge symmetries. The bosonic Polyakov path integral is a kind of 2d “quantum gravity” in the sense that we integrate over all possible 2d metrics g_{ab} . It is natural to look for a supersymmetric (SUSY) extension³⁹ of the construction. This leads to the *superstring*. The basic principles of the *Neveu–Schwarz/Ramond* (NS-R) approach

³⁹ We stress that history went the other way around: people formulated superstring theory, then realized that its world-sheet theory enjoyed a new kind of “symmetry” which was later generalized to 4d [73], and eventually called *supersymmetry*.

to superstring theory are *super*-reparametrization and *super*-Weyl invariance of the *super*-world-sheet.⁴⁰

Note 1.7 There are other (equivalent) approaches to the superstring, notably the *Green–Schwarz* one [74, 75] that may claim to be even more fundamental than the NS-R approach. However the covariant quantization becomes much harder in *Green–Schwarz* formulation, and concrete computations are typically less easy.

The world-sheet theory of the superstring is the supersymmetric version of the Polyakov action: it is a kind of 2d “supergravity” (SUGRA). In 2d we have *Majorana–Weyl* (MW) spinors⁴¹ with a single real component, and the 2d supersymmetry algebras are classified by two integers (p, q) where p is the number of supercharges which transform as chirality + MW spinors and q the number of chirality – MW supercharges; see Sect. 2.10 for more. The *superstring* corresponds to the minimal left–right symmetric choice, i.e. $(p, q) = (1, 1)$. The SUSY partner of the “dynamical” 2d metric g_{ab} is then a “dynamical” 2d Majorana gravitino χ_a . Neither field propagates local degrees of freedom in 2d, and the abusive adjective “dynamical” merely refers to the fact that these fields are integrated over in the path integral.

For a superstring propagating in flat space, the supercovariant action reads [76–78]

$$\begin{aligned} \frac{1}{4\pi} \int_{\Sigma} d^2\sigma \sqrt{g} \Big[& g^{ab} \partial_a X^\mu \partial_b X_\mu + i \psi^\mu \not{D} \psi_\mu + \\ & + i (\chi_a \gamma^b \gamma^a \psi^\mu) (\partial_b X_\mu - \frac{i}{4} \chi_b \psi_\mu) \Big] \end{aligned} \quad (1.127)$$

which is invariant under the local supersymmetry

$$\delta g_{ab} = 2i \epsilon \gamma_{(a} \chi_{b)}, \quad \delta \chi_a = 2 D_a \epsilon, \quad (1.128)$$

$$\delta X^\mu = i \epsilon \psi^\mu, \quad \delta \psi^\mu = \gamma^a (\partial_a X^\mu - \frac{i}{2} \chi_a \psi^\mu) \epsilon, \quad (1.129)$$

as well as under usual world-sheet reparametrizations. In Eq. (1.127) $\epsilon \equiv \epsilon(\sigma)$ is a coordinate-dependent Majorana spinor parameter, and the spacetime index μ takes the values $0, 1, \dots, d-1$. The action is invariant under Poincaré symmetry in the target space $\mathbb{R}^{d-1,1}$ with the “matter” fermions ψ^μ transforming as spacetime vectors. *Locally* on the world-sheet Σ we fix the 2d SUGRA (super)conformal gauge

$$g_{ab} = \rho \delta_{ab}, \quad \chi_a = \gamma_a \zeta, \quad (1.130)$$

where $\rho \equiv e^{2\phi}$ and ζ are the bosonic and fermionic component fields of the $(1, 1)$ Liouville superfield. Provided the (*super*)Weyl anomaly vanishes, the Liouville fields

⁴⁰ The world-sheet of the NS-R superstring is a *super*-manifold of complex dimension $1|1$.

⁴¹ We recall that a spinor χ is *Majorana* if it is real; $\chi^\dagger = \chi$ when written in a suitable representation of the γ -matrices. The spinor is *Weyl* if $\gamma_{d+1} \chi = \pm \chi$ where $\gamma_{d+1} = i^{(d(d-1)+2)/2} \gamma_0 \gamma_1 \dots \gamma_{d-1}$ is the chirality matrix. A *Majorana–Weyl* (MW) spinor satisfies both conditions. A pair of MW of opposite chirality forms a full Majorana fermion. See Sect. 8.1 for more details.

ρ, ζ drop out of the functional measure as in the bosonic theory, and the gauge-fixed version of the locally supersymmetric action (1.127) reduces to a free superconformal field theory (SCFT) plus gauge-fixing and ghost terms.

Quantization involves three issues:

- (1) as in the bosonic case, the requirement of no Weyl anomaly fixes the critical spacetime dimension of the theory d_{crit} ;
- (2) in addition to the “matter” fields ψ^μ, X^μ we have superconformal Faddeev–Popov ghosts with opposite statistics with respect to the corresponding gauge parameters: they are anticommuting (fermionic) for reparametrization symmetry and commuting (bosonic) for local supersymmetry. With the ghost fields we may perform a quantization *à la* BRST [22–25] of the superstring; this is the most natural and “modern” way to construct its physical states;
- (3) the superconformal structure of the super-world-sheet is non-trivial at the global level. Hence we end up with an integral over the finite-dimensional supermoduli $\mathcal{SM}_{g,n}$ of superconformal structures which are the SUSY counterpart to the complex moduli $\mathcal{M}_{g,n}$ for the bosonic string.

The equations of motion for the gauge fields g^{ab} and χ^a are the constraints

$$T_{ab} = S_a = 0, \quad (1.131)$$

where T_{ab} (resp. S_a) is the 2d energy–momentum tensor (resp. *supercurrent*). In the gauge-fixed theory these constraints will be reformulated in terms of BRST cohomology as usual; see Chap. 3.

Spin Structures on Σ

The world-sheet theory of the superstring contains 2d spinors. To define a theory with spinors on a manifold Σ , the manifold should be endowed with a *spin structure*; see BOX 1.10 for the basic facts in 2d. In general on a Riemann surface, there are several inequivalent spin structures [79, 80], in fact 2^{2g} of them. This can be understood as follows: on a genus g surface Σ , there are $2g$ non-contractible loops γ_i whose homotopy classes generate $H_1(\Sigma, \mathbb{Z}) \simeq \mathbb{Z}^{2g}$; a Majorana–Weyl fermion ψ may satisfy either the periodic or the anti-periodic b.c. along each basic loop γ_i , so the total number of possible choices of boundary conditions is 2^{2g} .

Residual Gauge Symmetry

We already know that the residual Diff^+ gauge symmetry, not fixed by the first condition in Eq. (1.130), are the motions generated by conformal Killing vectors v which satisfy the equation⁴²

$$\mathcal{L}_v g_{ab} = \sigma g_{ab} \quad \text{for some function } \sigma. \quad (1.132)$$

⁴² In 2d the CKV are simply the holomorphic vector fields; cf. Sect. 1.6.

BOX 1.10 - Spin structures on Riemann surfaces [79]

Roughly speaking, a 2d Weyl spinor ψ transforms as a “holomorphic $\frac{1}{2}$ -differential” $\psi'(dz')^{1/2} \approx \psi(dz)^{1/2}$ so that a bilinear in the fermions transforms as a covariant vector, i.e. as a differential $f dz$. Since a $(1,0)$ -form $f dz$ on a Riemann surface is a section of the canonical bundle $K \rightarrow \Sigma$, the fermions should be sections of line bundles $\mathcal{L} \rightarrow \Sigma$, with $\mathcal{L}^2 = K$. Note that there are no topological obstructions to the definition of “square-roots” of the canonical bundle, since the degree of K is always even, $\deg K = 2g - 2$, and hence $\deg \mathcal{L} = g - 1$ is an integer. A line bundle $\mathcal{L}$ such that $\mathcal{L}^2 = K$ is called a *spin structure*. We ask how many spin structures there are. Let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two spin bundles: $\mathcal{L}_1 \mathcal{L}_2^{-1}$ is a degree 0 line bundle whose square is trivial. Going back to BOX 1.6, we see that the line bundle $\mathcal{L}_1 \mathcal{L}_2^{-1}$ is a 2-torsion point in the Abelian group $J(\Sigma)$, i.e. and element $\xi \in J(\Sigma) \cong \mathbb{C}^g / H^1(\Sigma, \mathbb{Z})$ such that $2\xi = 0$, that is, ξ belongs to

$$\frac{1}{2} H^1(\Sigma, \mathbb{Z}) / H^1(\Sigma, \mathbb{Z}) \simeq \mathbb{Z}_2^{\text{rank } H^1(\Sigma, \mathbb{Z})} \equiv \mathbb{Z}_2^{2g}.$$

Thus on a genus g surface we have 2^{2g} distinct spin structures.

Physicists’ viewpoint In the physicists’ language, on Σ we have $2g$ independent cycles and a Weyl fermion may satisfy either the periodic or the anti-periodic b.c. along each of them, so that the total number of possible choices of boundary conditions is 2^{2g} .

Even and odd spin structures We distinguish the spin structures $\mathcal{L}$ into *even* and *odd* depending on the number mod 2 of zero-modes of the Weyl–Dirac operator $\bar{\partial}$ acting on sections of $\mathcal{L}$. Choose a symplectic basis of one cycle on Σ , i.e. $(A_i, B^i) \in H_1(\Sigma, \mathbb{Z})$ with $A_i \cdot A_j = B^i \cdot B^j = 0$ and $A_i \cdot B^j = \delta_i^j$ ($\cdot$ stands for the skew-symmetric intersection pairing in homology). Then a spin structure may be identified with an element of $H_1(\Sigma, \mathbb{Q})$ of the form

$$\frac{1}{2} \sum_{i=1}^g (a^i A_i + b_i B^i) \quad \text{with } a^i, b_i \in \{0, 1\}.$$

Theorem 1.7 [cf. [79]] *A spin structure is even iff $a^i b_i = 0 \bmod 2$.*

The number of even spin structures is $2^{g-1}(2^g + 1)$, while the number of odd ones is $2^{g-1}(2^g - 1)$.

Likewise, the local supersymmetries which leave the second condition (1.130) invariant are generated by spinorial parameters $\epsilon(z, \bar{z})$ which solve the equation⁴³

$$D_a \epsilon = \gamma_a \eta \quad \text{for some spinor } \eta. \quad (1.133)$$

A non-zero solution ϵ to (1.133) is called a *conformal Killing spinor* (CKS) [81–84]. The CKS generate the conformal *supersymmetries* of the super-Riemann surface. In the 2d Dirac matrix representation where

⁴³ D_a stands for the covariant derivative with respect to the world-sheet spin-connection $\omega_{a ij}$, that is, $D_a = \partial_a + \frac{1}{4} \omega_{a ij} \gamma^{ij}$, where $\gamma^{ij} = \gamma^i \gamma^j - \delta^{ij}$ are the generators of the spin group. γ^i are the 2d Dirac matrices, and $\gamma_a \equiv e_a^i \gamma_i$ with e_a^i the vielbein for the 2d metric g_{ab} .

$$\gamma_z = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \gamma_{\bar{z}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon \\ \tilde{\epsilon} \end{pmatrix}, \quad (1.134)$$

Eq. (1.133) becomes

$$\partial_{\bar{z}}\epsilon = \partial_z\tilde{\epsilon} = 0, \quad (1.135)$$

with ϵ (resp. $\tilde{\epsilon}$) a section of a holomorphic (resp. anti-holomorphic) line bundle which is a “square-root” of the holomorphic tangent bundle, that is, the *dual* $\mathcal{L}^{-1}$ of a spin bundle $\mathcal{L}$ such that

$$(\mathcal{L}^{-1})^2 = K^{-1} = T\Sigma; \quad (1.136)$$

see **BOX 1.10**. Thus a conformal Killing spinor is simply a holomorphic section $\epsilon(z)$ of the inverse of a spin bundle. The square $\epsilon(z)^2$ of a (commuting) CKS is a holomorphic vector field, hence a CKV: this is the dual relation to the SUSY algebra which states that the anticommutator of two superconformal supercharges is the generator of a conformal motion.

On a surface of genus $g \geq 2$ there are no CKV, hence there are *no* CKS either. On a torus⁴⁴ ($g = 1$) by definition there is *one* CKS if the spin structure $\mathcal{L}$ is *odd*, i.e. trivial $\mathcal{L} = \mathcal{O}$, and *none* if the spin structure is *even*, i.e. $\mathcal{L} \neq \mathcal{O}$. The complex dimension of the CKS space on the sphere follows from the Riemann–Roch theorem: $\dim_{\mathbb{C}} H^0(S^2, \mathcal{L}^{-1}) \equiv 1 - \frac{1}{2} \deg K = 2$, see **BOX 1.11** for the explicit CKS on S^2 .

Faddeev–Popov Ghosts for the Superstring

The FP ghosts have the same quantum numbers as the parameters of the corresponding gauge symmetry and opposite statistics. Hence they are (i) a complex Fermi vector field, $c \partial_z + \tilde{c} \partial_{\bar{z}}$, for reparametrization invariance as in the bosonic string, and (ii) a complex *Bose* 2d spinor

$$\begin{pmatrix} \gamma \\ \tilde{\gamma} \end{pmatrix}. \quad (1.137)$$

The anti-ghosts have the same statistics as their dual ghosts and the quantum numbers of the gauge-fixing conditions, which in the superconformal gauge (1.130) read

$$g_{zz} = g_{\bar{z}\bar{z}} = 0 \quad \text{and} \quad 2\chi_a - \gamma_a \gamma^b \chi_b = 0. \quad (1.138)$$

Thus we have fermionic quadratic differentials $b \equiv b_{zz}$ and $\tilde{b} \equiv \tilde{b}_{\bar{z}\bar{z}}$ (as in the bosonic string) and *bosonic* $\frac{3}{2}$ -differentials β and $\tilde{\beta}$. In conclusion, the ghost action is

$$S_{\text{ghost}} = \frac{1}{2\pi} \int_{\Sigma} \left(b \bar{\partial} c + \beta \bar{\partial} \gamma + \tilde{b} \partial \tilde{c} + \tilde{\beta} \partial \tilde{\gamma} \right) \quad (1.139)$$

⁴⁴ Note that on a torus for all spin bundle $\mathcal{L}$, we have $\mathcal{L} = \mathcal{L}^{-1}$ since $\mathcal{L}^2 = K \simeq \mathcal{O}$ because the torus has trivial canonical bundle.

BOX 1.11 - Superconformal symmetries of the 2-sphere

We identify the sphere with $\mathbb{P}^1(\mathbb{C})$ which is covered by two coordinate patches of coordinates z and w with $w = 1/z$. A CKV v is a holomorphic vector which has the expressions $f(z)\partial_z$, $f'(w)\partial_w$ in the two patches with $f(z)$, $f'(w)$ holomorphic. Agreement of the two expressions in the overlap of the two patches requires

$$f'(w) = f(z) \frac{\partial w}{\partial z} \Big|_{z=1/w} = -f(1/w) w^2.$$

Then $f(z)$ has a pole of degree at most 2 at infinity, i.e.

$$v(z) = (a + bz + cz^2)\partial_z \quad (\text{CKV})$$

for a, b, c complex constants, so the space of CKV has complex dimension 3. The same analysis for the CKS yields

$$\epsilon'(w) = i\epsilon(1/w)w$$

and $\epsilon = (\alpha + \beta z)\kappa$ with α, β constants and $\kappa^2 = \partial_z$, so the space of CKS has complex dimension 2. Setting

$$J_- = \partial_z, \quad J_0 = z\partial_z, \quad J_+ = z^2\partial_z, \quad Q = \kappa, \quad S = z\kappa,$$

we have the superconformal super-Lie algebra of the sphere

$$[J_0, J_{\pm}] = \pm J_{\pm}, \quad [J_-, J_+] = 2J_0, \quad \{Q, Q\} = 2J_-, \quad \{Q, S\} = 2J_0, \quad \{S, S\} = 2J_+.$$

In Kac's classification of simple super-Lie groups [85, 86], this corresponds to the complex $OSp(1|2)$ whose bosonic subgroup is $O(1) \times Sp(2, \mathbb{C}) \cong SL(2, \mathbb{C})$ and the fermionic generators form a single copy of the fundamental.

while—assuming the super-Weyl anomaly cancels—in the superconformal gauge the matter part of the action reduces to the free action

$$S_{\text{matter}} = \frac{1}{2\pi} \int_{\Sigma} \left(\partial X^{\mu} \bar{\partial} X_{\mu} + \psi^{\mu} \bar{\partial} \psi_{\mu} + \tilde{\psi}^{\mu} \partial \tilde{\psi}_{\mu} \right) \quad (1.140)$$

($\mu = 0, 1, \dots, d-1$). We stress that all 2d actions are the straightforward supersymmetrization of their bosonic counterparts.

As in the bosonic string, the critical dimension d_{crit} is fixed by requiring the super-Weyl anomaly to vanish. Since supersymmetry is non-anomalous, the numerical coefficients in front of the bosonic and fermionic parts of the anomaly are equal, and the full super-Weyl anomaly vanishes iff the usual bosonic Weyl anomaly cancels. All 2d fields⁴⁵ are free with a Dirac-like action (1.94) and spin λ as in Table 1.3. We compute the Weyl anomaly from Eq. (1.95): the coefficient of R in $-12 T^a_a$ is

⁴⁵ We consider only the left-movers; the story for the right-movers is identical.

Table 1.3 World-sheet fields entering in the conformal gauge superstring action

Field	Statistics	Reality	λ	$-(-1)^F(1 - 3(1 - 2\lambda)^2)$
X^μ	Bosonic	Real	0	+2
ψ^μ	Fermionic	Real	$\frac{1}{2}$	+1
b, c	Fermionic	Complex	2	-26
β, γ	Bosonic	Complex	$\frac{3}{2}$	+11

$$\overbrace{\frac{d}{2} \cdot (2 + 1)}^{\text{matter}} \quad \overbrace{- 26 + 11}^{\text{ghosts}} \equiv \frac{3d - 30}{2}, \quad (1.141)$$

so the critical dimension of the superstring is $d_{\text{crit}} = 10$.⁴⁶

Ghost Zero-Modes

As in the bosonic case, the zero-modes of the ghosts c, γ correspond to the residual gauge symmetries, while the zero-modes of the anti-ghosts b, β to deformations of the metric and gravitino fields which are not gauge transformations, i.e. the b, β zero-modes span the tangent space to the *supermoduli* of the super-Riemann surface. Supermoduli are very complicated superspaces [87–91] except in genus zero (where they are trivial) and in genus $g = 1$ and 2 where they split in the complex moduli of the underlying bosonic surface and the odd moduli.

Again, ghost zero-modes produce finite-dimensional Faddeev–Popov determinants. For b, c zero-modes the story is as in the bosonic string. For β, γ the situation is much more involved, since these fields are *bosonic*, and we cannot “soak up” their zero-modes by simply inserting fields in the path integral. Indeed, for free fermions the path integral is a determinant and a zero-eigenvalue makes the determinant zero, whereas for free bosons the path integral is the *inverse* of a determinant, and a zero-eigenvalue produces a naively divergent answer.

To define the path integral measure in the SUSY case is very subtle, and requires advanced conformal theory techniques that we shall develop in the next chapter and further study in Chap. 10.

1.8 Strings Moving in Curved Backgrounds

Up to now, we limited ourselves to (super)strings moving in a *flat* d -dimensional spacetime M . However we saw in Sect. 1.3.2 that the target-space physics contains dynamical gravity, and hence the geometry of spacetime is dynamical and cannot be

⁴⁶ Cf. the “prediction” from the classification of the hyperKähler manifolds in **BOX 1.3**.

fixed a priori. We are forced to consider strings propagating in a curved background [92–94]. The obvious generalization of the bosonic Polyakov action is

$$S = \int_{\Sigma} \frac{d^2\sigma}{2\pi\alpha'} \sqrt{g} \left(\frac{1}{2} g^{ab} G(X)_{\mu\nu} \partial_a X^\mu \partial_b X^\nu + \dots \right) \quad (1.142)$$

where $G(X)_{\mu\nu}$ is the background target metric, and the 2d fields X^μ now are *local* coordinates in the spacetime M . Equation (1.142) is the action⁴⁷ of the 2d *non-linear* σ -model with *target space* M . The field configurations $X^\mu(\sigma)$ of the σ -model are invariantly seen as maps between the Riemannian manifolds (Σ, g_{ab}) and $(M, G_{\mu\nu})$

$$X: \Sigma \rightarrow M. \quad (1.143)$$

The classical Euclidean solutions are (by definition) the *harmonic maps* [95, 96].

The background metric $G(X)_{\mu\nu}$ should be really seen as the v.e.v. $\langle G(X)_{\mu\nu} \rangle$ of the dynamical massless field $G(X)_{\mu\nu}$ whose on-shell states we discovered in the light-cone gauge: they are described by the transverse states

$$(\alpha_i \tilde{\alpha}_j + \alpha_j \tilde{\alpha}_i)_{\text{traceless}} |0, p_\mu\rangle. \quad (1.144)$$

More generally, we may have non-trivial backgrounds for the target-space fields corresponding to all (infinitely many) on-shell states of the string. Each of these background fields corresponds to an operator we may add to the 2d action on the world-sheet. Massive states correspond to operators containing more than two derivatives, the tachyon to a no-derivative potential term, and the massless field background to 2-derivative terms in the action. We are particularly interested in the massless field backgrounds. In the bosonic string we have three such fields: the metric $G_{\mu\nu}$, the B -field 2-form $B_{\mu\nu}$, and the dilaton Φ . The corresponding world-sheet (Euclidean) action is [92–94]

$$\frac{1}{4\pi\alpha'} \int_{\Sigma} \sqrt{g} d^2\sigma \left\{ (g^{ab} G(X)_{\mu\nu} + i\epsilon^{ab} B(X)_{\mu\nu}) \partial_a X^\mu \partial_b X^\nu + \alpha' R \Phi(X) \right\}. \quad (1.145)$$

This action is manifestly invariant under arbitrary 2d field redefinitions

$$X^\mu \rightarrow X'^\mu = X'^\mu(X^\nu) \quad (1.146)$$

which yield target-space reparametrizations. It is also invariant under B -field gauge transformations

$$B \rightarrow B + d\eta, \quad \eta \text{ an arbitrary 1-form.} \quad (1.147)$$

⁴⁷ The σ -model action is known as the *Dirichlet integral* in the math literature; see, for example, [95].

The quickest way to check this fact is to see the 2d fields as maps $X: \Sigma \rightarrow M$, so the B -field part of the action may be written as the integral of the pull-back [97]⁴⁸ of B

$$\frac{i}{2\pi\alpha'} \int_{\Sigma} X^* B \quad (1.148)$$

whose gauge variation is

$$\delta \left(\frac{i}{2\pi\alpha'} \int_{\Sigma} X^* B \right) = \frac{i}{2\pi\alpha'} \int_{\Sigma} d(X^* \eta) \equiv \frac{i}{2\pi\alpha'} \int_{\partial\Sigma} X^* \eta \quad (1.149)$$

which vanishes for a closed world-sheet ($\partial\Sigma = \emptyset$). In the presence of a boundary the variation (1.149) does not vanish, and this leads to a mixing of the B -field gauge symmetry (1.147) with other local symmetries: this is an important aspect of open string theory on which we shall return later in the book.

The gauge-invariant field strength of the B -field is the 3-form

$$H = dB. \quad (1.150)$$

The physics of the 2d model (1.145) depends only on the gauge-invariant objects that we may construct out of the metric and the other background fields.

The world-sheet theory in a non-trivial background is an *interacting* 2d QFT. We see the metric, B -field, and dilaton backgrounds as couplings of this QFT. Expanding the background fields around a constant classical solution x_0^μ in powers of the quantum fluctuation field $X_q^\mu \equiv X^\mu - x_0^\mu$, we see that each target-space field actually combines *infinitely many* 2d couplings. For example, the metric has a Taylor expansion in normal coordinates [98, 99] of the form

$$\begin{aligned} G_{\mu\nu}(X) &= \sum_{n \geq 0} G^{(n)}(x_0)_{\mu\nu\rho_1 \dots \rho_n} X_q^{\rho_1} \dots X_q^{\rho_n} = \\ &= \delta_{\mu\nu} - \frac{1}{3} R_{\mu\rho_1\nu\rho_2} X_q^{\rho_1} X_q^{\rho_2} + \frac{1}{6} \nabla_{\rho_1} R_{\mu\rho_2\rho_3\nu} X_q^{\rho_1} X_q^{\rho_2} X_q^{\rho_3} + \dots \end{aligned} \quad (1.151)$$

whose coefficients $G^{(n)}(x_0)_{\mu\nu\rho_1 \dots \rho_n}$ are *universal* polynomials in the covariant derivatives of the curvature tensor at the point $x_0 \in M$. Inserting this expansion in (1.145), we see that the n -th coefficient $G^{(n)}(x_0)_{\mu\nu\rho_1 \dots \rho_n}$ is the couplings for the 2d quantum operator $X_q^{\rho_1} \dots X_q^{\rho_n} \partial^a X_q^\mu \partial_a X_q^\nu$.

In a background which is non-trivial *only* for the massless fields, the 2d QFT (1.145) is *renormalizable*, so we can use standard QFT techniques to study it.

Note 1.8 Comparing (1.145) with Eq. (1.12) we see that the Euler term in the Polyakov action corresponds to a constant background for the dilaton field, i.e. $\lambda = \langle \Phi \rangle$. Hence in string theory the coupling e^λ is not a numerical parameter, but the vacuum expectation value of the dynamical field Φ . String theory does not

⁴⁸ The i in front of the B -term in the Euclidean action makes the Minkowski action real.

contain any adjustable parameter—*i.e. string theory is really unique*. The absence of adjustable parameters is believed to be a general feature of *all* consistent theories of quantum gravity as argued in the **Introduction**; see also [100].

Weyl Invariance in Non-trivial Backgrounds

The trace of energy–momentum tensor measures the non-invariance under Weyl transformations; cf. (1.91). Classically, the only term in the 2-derivative action (1.145) which breaks Weyl invariance is the dilaton coupling, but quantum mechanically the story is subtler: in a renormalizable QFT the trace T^a_a is specified by the β -functions of the various couplings [101]. For the 2d model (1.145) we have

$$T^a_a = -\frac{1}{2\alpha'} \beta_{\mu\nu}^G g^{ab} \partial_a X^\mu \partial_b X^\nu - \frac{i}{2\alpha'} \beta_{\mu\nu}^B \epsilon^{ab} \partial_a X^\mu \partial_b X^\nu - \frac{1}{2} \beta^\Phi R \quad (1.152)$$

where, say, $\beta_{\mu\nu}^G g^{ab} \partial_a X^\mu \partial_b X^\nu$ actually stands for the infinitely many couplings

$$\sum_{n \geq 0} \beta_{\mu\nu\rho_1 \dots \rho_n}^{(n)} X_q^{\rho_1} \dots X_q^{\rho_n} g^{ab} \partial_a X^\mu \partial_b X^\nu \quad (1.153)$$

with $\beta_{\mu\nu\rho_1 \dots \rho_n}^{(n)}$ the β -function of the coupling $G_{\mu\nu\rho_1 \dots \rho_n}^{(n)}$.

As always in perturbative QFT, we may expand the β -functions into contributions from different loop orders. The ℓ -loop contributions to $\beta_{\mu\nu}^G$, $\beta_{\mu\nu}^B$, and β^Φ are again covariant tensors in target-space given by *universal* polynomials in the covariant derivatives of the “curvatures” $R_{\mu\nu\rho\sigma}$, $H_{\mu\nu\rho}$, and $\nabla_\mu \Phi$. Standard Feynmann diagram combinatorics shows that the ℓ -loop contribution scales as⁴⁹ $(\alpha')^\ell$. Since α' has the dimension $(\text{length})^2$, the loop expansion is also the expansion in the number of spacetime derivatives: *the ℓ -loop term contains 2ℓ derivatives*. The Ricci identity [28]

$$[D_\mu, D_\nu]^i_j = R_{\mu\nu}^i_j \quad (1.154)$$

implies that a Riemann tensor counts for two derivatives while $H_{\mu\nu\rho}$ and $D_\mu \Phi$ count for one. In conclusion: ***the world-sheet loop expansion (a.k.a. α' -expansion) is a low-momenta series which produces the spacetime low-energy effective action.*** Clearly this expansion is reliable, and the effective action meaningful, when the spacetime curvatures (and their derivatives) are small in string units $\alpha' |R_{\mu\nu\rho\sigma}| \ll 1$.

We give the expansions of the β -functions in Eq. (1.152) up to two spacetime derivatives.⁵⁰ Tree-level terms correspond to the free-field anomaly proportional to $d - 26$ discussed before (the -26 comes from the free ghosts b , c which do not couple to the spacetime geometry). All other contributions are of order $O(\alpha')$ or higher:

⁴⁹ The overall factor $1/(2\pi\alpha')$ in (1.145) plays the role of $1/\hbar$ in front of the action, where $\hbar$ is the loop-counting parameter.

⁵⁰ For simplicity we consider the bosonic string. The case of the superstring involves additional issues which will be discussed later in the book. For more details on the β -functions, see [92–94].

$$\beta_{\mu\nu}^G = \alpha' R_{\mu\nu} + 2\alpha' D_\mu D_\nu \Phi - \frac{\alpha'}{4} H_{\mu\rho\sigma} H_\nu^{\rho\sigma} + O(\alpha'^2) \quad (1.155)$$

$$\beta_{\mu\nu}^B = -\frac{\alpha'}{2} D^\rho H_{\rho\mu\nu} + \alpha' (D^\rho \Phi) H_{\rho\mu\nu} + O(\alpha'^2) \quad (1.156)$$

$$\beta^\Phi = \frac{d-26}{6} - \frac{\alpha'}{2} D^2 \Phi + \alpha' D_\rho \Phi D^\rho \Phi - \frac{\alpha'}{24} H_{\mu\nu\rho} H^{\mu\nu\rho} + O(\alpha'^2). \quad (1.157)$$

As advertised, these β -functions are covariant tensors under target-space diffeomorphisms and B -field gauge transformations (1.147). They contain the dilaton Φ only through its derivatives $D_\mu \Phi$, since a constant value of Φ corresponds to the topological term λR in the Polyakov action (1.12) which does not contribute to the local dynamics in two-dimensions and hence does not spoil conformal invariance.

The world-sheet action (1.145) defines a consistent bosonic string theory (moving in a non-trivial background) if and only if it satisfies the condition of Weyl invariance, $T^a_a = 0$. From Eq. (1.152), $T^a_a = 0$ is equivalent to the set of equations

$$\beta_{\mu\nu}^G = \beta_{\mu\nu}^B = \beta^\Phi = 0. \quad (1.158)$$

These equations have a simple physical interpretation: they are the *field equations* satisfied by the spacetime massless fields $G_{\mu\nu}$, $B_{\mu\nu}$, and Φ . That is, ***a string propagates consistently in a spacetime background if and only if the background is on-shell, that is, iff it satisfies the field equations*** (1.158).

To understand the physical meaning of this statement, we look at each equation in turn. In view of Eq. (1.155), the condition $\beta_{\mu\nu}^G = 0$ may be rewritten in the form

$$\left(R_{\mu\nu} - \frac{1}{2} G_{\mu\nu} R \right) = \frac{1}{4} \left(H_{\mu\rho\sigma} H_\nu^{\rho\sigma} - \frac{1}{2} G_{\mu\nu} H_{\rho\sigma\tau} H^{\rho\sigma\tau} \right) + \dots \quad (1.159)$$

which has the form of the Einstein equations for the spacetime metric $G_{\mu\nu}$ with a certain “matter” energy–momentum tensor, namely the RHS of (1.159). Likewise, $\beta_{\mu\nu}^B = 0$ has the typical form of an equation of motion for a gauge 2-form where the divergence of its field strength H is set equal to some source “current”:

$$d * dB = \text{“source terms”}, \quad (1.160)$$

and the same holds for the third equation which has the typical form of the scalar field equation:

$$- D^a D_a \Phi = \text{“source terms”}. \quad (1.161)$$

An even more convincing proof that this is the correct physical interpretation of the world-sheet β -functions comes from the realization that the above equations can be obtained varying an effective action in spacetime which has all the physically required properties; see Sect. 1.8.1 below.

The Linear Dilaton

The critical dimension condition for flat backgrounds, $d = 26$, gets replaced in a non-trivial background by the more general equation

$$\beta^\Phi = 0. \quad (1.162)$$

An *exact* solution to this condition, which also solves the other two equations $\beta^G = \beta^B = 0$, is given by the *linear dilaton*

$$G_{\mu\nu} = \eta_{\mu\nu}, \quad B_{\mu\nu} = 0, \quad \Phi = V_\mu X^\mu, \quad V_\mu V^\mu = \frac{26-d}{6\alpha'} \quad (1.163)$$

with V_μ a constant vector. In this background the world-sheet action (1.145) is still Gaussian, hence we may compute exactly all quantum amplitudes, and check the Weyl invariance explicitly; see next chapter. Equation (1.163) is an exact on-shell background for the bosonic string moving in $d < 26$ dimensions, but the d -dimensional Lorentz invariance is broken by the constant gradient $D_\mu \Phi \equiv V_\mu$ of the dilaton.

1.8.1 The Spacetime Effective Action

We leave as an exercise for the reader to show that the equations of motion (1.155)–(1.157) may be obtained from the spacetime action

$$S = \frac{1}{2\kappa_0^2} \int d^d x \sqrt{-G} e^{-2\Phi} \left[-\frac{2(d-26)}{3\alpha'} + R - \frac{1}{12} H_{\mu\nu\rho} H^{\mu\nu\rho} + 4\partial_\mu \Phi \partial^\mu \Phi + O(\alpha') \right]. \quad (1.164)$$

The constant κ_0^2 has no physical significance: it can be set to any chosen positive value by a suitable field redefinition of the form

$$\Phi \rightarrow \Phi + \text{const.} \quad (1.165)$$

since the effective Lagrangian has the overall factor $e^{-2\Phi}$. Except for this overall factor, the dilaton Φ enters the *tree-level* effective Lagrangian only through its derivatives, so a shift of Φ by a constant has the effect of a rigid overall rescaling of the action. The factor $e^{-2\Phi}$ is present because the tree-level effective Lagrangian describes string amplitudes on a world-sheet with $g = 0$, i.e. on the sphere S^2 : the overall factor $e^{-2\Phi} \equiv e^{-\langle \Phi \rangle \chi(S^2)}$ just reflects the topological term in the Polyakov action (1.12).

String Frame Versus Einstein Frame

The above action (1.164) is written in the so-called *string frame* meaning that the target-space fields $G_{\mu\nu}$, $B_{\mu\nu}$, and Φ are the ones appearing as couplings in the string world-sheet theory (1.145).

The fields used in the standard formulation of the target-space physics are related to these ones by a field redefinition. In the standard GR conventions, the fields are defined so that the gravity term in the action has the Einstein–Hilbert form

$$-\frac{1}{2\kappa^2} \int \sqrt{-g} R, \quad (1.166)$$

without extra field-dependent factors in front of the scalar curvature R . The fields of standard target-space formulation are said to be in the *Einstein frame*.

To pass from the string frame to the Einstein one, we perform a Weyl redefinition of the metric: see **BOX 1.12** for the Weyl transformations of various quantities in d dimensions. We write $G_{\mu\nu}^E$ for the Einstein frame metric

BOX 1.12 - Conformal properties of $R_{\mu\nu}$ and R in d dimensions

We quote (a small part of) **Theorem 1.159** of [102]:

Under the replacement $g_{\mu\nu} \rightarrow e^{2f} g_{\mu\nu}$ one has

$$\begin{aligned} R_{\mu\nu} &\rightarrow R_{\mu\nu} - (d-2)(D_\mu \partial_\nu f - \partial_\mu f \partial_\nu f) - (D^\mu \partial_\mu f + (d-2)|\partial f|^2)g_{\mu\nu} \\ R &\rightarrow e^{-2f} \left(R - 2(d-1)D^\mu \partial_\mu f - (d-2)(d-1)|df|^2 \right). \end{aligned}$$

In particular

$$\sqrt{-g}R \rightarrow e^{(d-2)f} \sqrt{-g} \left(R - 2(d-1)D^\mu \partial_\mu f - (d-2)(d-1)|df|^2 \right). \quad (\spadesuit)$$

$$G_{\mu\nu}^E = \exp[-4\Phi/(d-2)] G_{\mu\nu}. \quad (1.167)$$

Using the formulae in **BOX 1.12**, we get

$$\begin{aligned} S^E = \frac{1}{2\kappa^2} \int d^d x \sqrt{-G^E} \Bigg[&-\frac{2(d-26)}{3\alpha'} e^{4\Phi/(d-2)} + R - \\ &-\frac{1}{12} e^{-8\Phi/(d-2)} H_{\mu\nu\rho} H^{\mu\nu\rho} - \frac{4}{d-2} \partial_\mu \Phi \partial^\mu \Phi + O(\alpha') \Bigg]. \end{aligned} \quad (1.168)$$

1.8.2 String Compactifications

One apparently non-realistic feature of string theory is that the number of spacetime dimensions of a critical string ($d = 26$ for the bosonic, $d = 10$ for the superstring) is larger than the observed 4 macroscopic dimensions. This is not necessarily a problem: the spacetime metric $G_{\mu\nu}$ is a dynamical field, so the theory contains sectors describing the motion of the string in target spaces of different topologies and geometries. In particular we have spacetimes of the form

$$\mathbb{R}^{n-1,1} \times K_{d-n}, \quad n < d \quad (1.169)$$

where $\mathbb{R}^{n-1,1}$ is the n -dimensional Minkowski space and K_{d-n} is a $(d-n)$ -dimensional compact manifold whose diameter is small enough so that the $(d-n)$ internal dimensions are not detectable in low-energy experiments.⁵¹

Writing X^μ for the coordinates in $\mathbb{R}^{n-1,1}$ and Y^i for local coordinates in K_{d-n} , the world-sheet theory for such a curved on-shell background takes the form

$$\frac{1}{4\pi\alpha'} \int \sqrt{g} d^2\sigma \left(\overbrace{g^{ab} \partial_a X^\mu \partial_b X_\mu}^{\text{Poincare invariant } n\text{-dimensional}} + \overbrace{g^{ab} k_{ij} \partial_a Y^i \partial_b Y^j + \dots}^{\text{internal CFT}} \right), \quad (1.170)$$

that is, the world-sheet theory decomposes in a free theory which describes the motion of the string in the “visible” flat factor $\mathbb{R}^{n-1,1}$, and an interacting “internal” 2d conformal field theory (CFT) given by the σ -model with target space the Riemannian manifold K_{d-n} endowed with a metric k_{ij} which satisfies the equation

$$\beta_{ij}^k = 0, \quad (1.171)$$

as required by Weyl invariance. At the leading order in an expansion in powers of α' , Eq. (1.171) says that the Riemannian manifold K_{d-n} must be *Ricci-flat*

$$R_{ij} + O(\alpha') = 0. \quad (1.172)$$

From this discussion, we learn that ***we can replace the “internal” σ -model in Eq. (1.170) with ANY 2d conformal field theory producing the right Weyl anomaly to cancel the contribution from the ghosts and the non-compact fields X^μ .***

The same story applies *mutatis mutandis* to the superstring. The world-sheet actions are the supersymmetric versions of the ones for the bosonic string, and the “internal” world-sheet theory may be any $(1, 1)$ superconformal field theory with the appropriate Weyl anomaly.

⁵¹ Luckily this statement is not really correct. The internal geometry is (in principle) detectable with experiments at low energy, but these experiments cannot involve *solely* localized wave-packets of “usual” particles. Internal dimensions are invisible to low-energy *phenomenologists*, not to low-energy *phenomena*. See, for example, [103].

In the supersymmetric case, we shall be more precise and look for *exact* solutions to the β -function equation (1.171) which describe supersymmetry preserving, Poincaré invariant, vacua of the superstring moving in $d < 10$ effective (i.e. non-compact) dimensions, including the “realistic” case $d = 4$: see Chap. 11. In BOX 1.13 you find a heuristic experimental motivation to believe that the extra dimensions are for real.

In conclusion, to understand the world-sheet theory of the string (resp. superstring) we need to understand the 2d conformal fields theories (resp. 2d superconformal QFT). We shall systematically study these theories in the next chapter.

1.9 Physical Amplitudes, S-Matrix, and Vertices

We have defined the string partition function (vacuum amplitude) via the Polyakov path integral and established (in principle) the e.o.m. for the background fields. This is not yet the full story: to construct a physical theory, we need to define *all* observables and compute the amplitudes for *all* physical processes (at least in principle).

BOX 1.13 - Extra dimensions: are they “real”? (Heuristics)

Some phenomenology guy thinks that superstring theory is not “realistic” because it predicts a large number of dimensions, $d = 10$, whereas we “observe” only 4. Is this correct? Is there no experimental indication for extra dimensions besides the macroscopic four?

Albeit at a moral level rather than as a technical statement there is indeed a huge experimental evidence that extra dimensions ought to exist. The story goes as follows: we have overwhelming experimental evidence for three statements:

- (i) in the real world, the speed of light is a constant c independent of the frame
- (ii) gravity exists
- (iii) the world is quantum, i.e. $\hbar \neq 0$.

The first two items imply that in the extreme IR, four-dimensional Einstein gravity theory should hold; then there exist black holes at least for large Schwarzschild radius R . From (iii) this implies an entropy $S \sim \pi R^2$ (as $R \rightarrow \infty$) and hence e^S quantum states. The no-hair theorem suggests that these quantum states cannot live in four-dimensions. Hence they must live in the extra dimensions. In facts this is the way it works in the example of consistent quantum gravity given by the superstring (see Sect. 14.6). Hence—at least morally speaking—the three experimental facts (i), (ii), and (iii), taken together, provide “strong” evidence for extra dimensions.

n -Graviton Processes

In particle physics, the basic observables are the S -matrix elements, i.e. the on-shell amplitudes between incoming multi-particle states $|\text{in}\rangle$ at $t = -\infty$ and outgoing multi-particle states $|\text{out}\rangle$ at $t = +\infty$. When the theory contains dynamical gravity, the notion of S -matrix becomes tricky. In general it is not defined, unless our

spacetime M is asymptotic at infinity to flat Minkowski space [104, 105]. Our goal is to compute the S -matrix elements of string theory (in an asymptotically flat background) in terms of Polyakov path integrals. We shall return to this central problem several times in this book. Here we are interested in the schematic form of the answer, omitting all details. To simplify the exposition we focus on the bosonic string.

As an illustrative example, we consider a scattering process involving a number of gravitons. We may see the gravitons as small fluctuations around the background metric which we take to be the flat one $\eta_{\mu\nu}$,

$$G(X)_{\mu\nu} = \eta_{\mu\nu} + h(X)_{\mu\nu}. \quad (1.173)$$

The small fluctuation $h(X)_{\mu\nu}$ satisfies the linearized equations of motion. Together with the gauge-fixing conditions, the linearized Einstein equations read

$$\square h_{\mu\nu} = \partial^\mu h_{\mu\nu} = h^\mu{}_\mu = 0. \quad (1.174)$$

By general theory, $h_{\mu\nu}$ is a linear superposition of plane waves of the form

$$\epsilon_{\mu\nu} e^{ik \cdot X} \quad \text{where} \quad k^2 = k^\mu \epsilon_{\mu\nu} = \epsilon^\mu{}_\mu = 0, \quad \epsilon_{\mu\nu} = \epsilon_{\nu\mu}. \quad (1.175)$$

Two polarizations $\epsilon_{\mu\nu}, \epsilon'_{\mu\nu}$ are gauge equivalent iff $\epsilon'_{\mu\nu} - \epsilon_{\mu\nu} = k_{(\mu} v_{\nu)}$ with $k^\mu v_\mu = 0$.

At low momenta the Einstein equations reduce to the vanishing of the world-sheet β -functions; cf. Sect. 1.8. The small fluctuation $h_{\mu\nu}$ represents an *on-shell graviton*, iff it satisfies the linearized equations (1.174), i.e. iff the operator $h_{\mu\nu} \partial^a X^\mu \partial_a X^\nu$ is a *marginal deformation* of the free 2d action which preserves the Weyl invariance (to lowest order). Formally, the path integral in the deformed background is

$$\begin{aligned} & \sum_{n=0}^{\infty} \frac{1}{n!} \int [dX db dc d\tilde{b} d\tilde{c}] e^{-S_{\text{free}}} \left(-\frac{1}{2\pi} \int_{\Sigma} h(X)_{\mu\nu} \partial X^\mu \bar{\partial} X^\nu \right)^n = \\ & = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{g=0}^{\infty} e^{-\lambda(2-2g)} \int_{\text{genus } g} [dX db dc d\tilde{b} d\tilde{c}] e^{-S_{\text{free}}} \left(\int_{\Sigma} O(z) \right)^n \end{aligned} \quad (1.176)$$

where

$$O(z) \equiv -\frac{1}{2\pi} h(X)_{\mu\nu} \partial X^\mu \bar{\partial} X^\nu(z). \quad (1.177)$$

In view of (1.175) the n -th term in the first sum of (1.176) describes a process involving n on-shell gravitons. We claim that—when the expression is given its correct definition—the corresponding amplitude is an n -graviton S -matrix element. The rest of this section is dedicated to justify our claim. We begin with some general remarks on the structure of the Polyakov quantum amplitudes.

Fixing the Diff^+ Gauge

The summands in the second line of (1.176) are integrals over the positions of n insertions of the operator $O(z)$ in a world-sheet Σ of genus g . However these formal expressions are not really correct, because the Diff^+ gauge symmetry “moves points around” in Σ , and insertions of operators at different points may actually be gauge equivalent, so integrating over all insertion points may result in double-counting.

The intrinsic viewpoint is that each summand is a path integral over surfaces Σ of genus g with n punctures ($\equiv$ the points where we insert the operators $O(z)$). After fixing the **Weyl** and Diff^+ gauges, we remain with an integral over the finite-dimensional moduli space $\mathcal{M}_{g,n}$ of complex structures of genus g curves with n punctures. The integral over $\mathcal{M}_{g,n}$ *unifies* in a gauge-invariant way the integrals over the world-sheet metrics and over the positions of the n operator insertions.

Consider, say, $g = 0$ and $n = 3$: the sphere with three punctures is *rigid*, i.e. $\mathcal{M}_{0,3}$ is a point; cf. Eq. (1.119). Therefore the 3-point amplitudes may be computed by fixing the three operator insertions at any convenient (distinct) points on S^2 while inserting 3 c and 3 $\tilde{c}$ ghosts to “soak up” the ghost zero-modes, or more properly, to produce the finite-dimensional Faddeev–Popov determinant which yields the correct measure on $\mathcal{M}_{0,3}$. More generally,

$$\dim_{\mathbb{C}} \mathcal{M}_{0,n} = n - 3, \quad (1.178)$$

since we may put 3 out of the n insertions in any pre-assigned position with a suitable diffeomorphism. In practice, when computing n graviton amplitudes on the sphere we insert the operators

$$c(\hat{z}_i) \tilde{c}(\hat{\bar{z}}_i) O_i(\hat{z}_i) \quad i = 1, 2, 3, \quad (1.179)$$

at the 3 points $\hat{z}_i$ which are kept fixed as a residual gauge condition, cf. Eq. (1.104), while we integrate over the positions of the other $n - 3$ insertions

$$\int O_i(z_i) d^2 z_i \quad i = 4, 5, \dots, n. \quad (1.180)$$

The resulting amplitude is however perfectly symmetric between the n insertions by the very definition of $\mathcal{M}_{0,n}$. We shall return to the explicit form of the measure in Chaps. 3, 4, and 10.

String S -matrix

Let us give a physical explanation of why—after taking into account the correct measure on $\mathcal{M}_{g,n}$ —the n -th term in the sum in the first line of (1.176) is a n -graviton S -matrix element written as a perturbative sum of stringy g -loop contributions.⁵²

⁵² It is crucial not to confuse two very different notions of “perturbative” corrections. On the one hand, we have the *stringy* loop corrections: the g loop contribution is given by a 2d path integral over surfaces of genus g which may be interpreted as genuine quantum processes involving intermediate

We work at the tree level, i.e. $\Sigma = S^2$. Consider one non-integrated graviton vertex (1.179) inserted, say, at the point $\hat{z}_1 = 0$. The vicinity of this point in $\Sigma \equiv \mathbb{P}^1$ is biholomorphic to the open punctured disk $\Delta^* \equiv \{z \in \mathbb{C}: 0 < |z| < 1\}$. Writing

$$z = \exp(2\pi i w) \quad \text{with} \quad w \sim w + 1 \quad (1.181)$$

the punctured disk Δ^* is mapped biholomorphically ($\equiv$ conformally) to the flat half-cylinder C of infinite length

$$C = \{w \in \mathbb{C}, \operatorname{Im} w > (2\pi)^{-1}\} / (w \sim w + 1). \quad (1.182)$$

We see the coordinate $\tau = \operatorname{Im} w$ as the (Euclidean) time on C . As discussed at the beginning of this chapter, a cylinder represents the free evolution of the closed string. Wick rotating to the physical Minkowski 2d time, and using the *physical* light-cone gauge on the cylinder,

$$X^+ \propto \tau, \quad (1.183)$$

we see that infinite world-sheet time means infinite physical time (in target space). Then the marginal operator $\mathcal{O}$ inserted at $z = 0$ represents a small on-shell deformation $h_{\mu\nu}$ of the metric in the infinite future (or past), that is, it represents an asymptotic scattering state of the graviton particle. This conclusion extends to the integrated vertices (1.180) by virtue of **Diff**⁺ and **Weyl** gauge symmetries.

The physical process described by the above amplitude then corresponds to a transition between asymptotic states at $t = -\infty$ (for vertices whose momentum k_μ in Eq. (1.175) is in-coming) and asymptotic states at $t = +\infty$ (vertices with out-going momenta). World-sheet Weyl invariance requires these asymptotic states to satisfy the linearized equations of motion, that is, they are *on-shell asymptotic states*. Therefore **the n -th summand in (1.176) is the on-shell S-matrix element for a scattering process involving n gravitons** (counting in-going and out-going ones).

In string theory it is hard to define off-shell amplitudes, and we do not know in general how to compute them. However conformal amplitudes of the above form give a prescription to compute arbitrary on-shell scattering process (as a perturbative expansion). This prescription defines perturbative string theory. In Part IV of the book we shall consider string theory beyond its perturbative definition.

Physical Vertices

From the above discussion, it is clear that there is a direct relation between the asymptotic on-shell states and the exactly marginal local operators we may insert at

states with multiple virtual strings. These are the actual quantum correction as seen from the target-space physics. The *stringy* loop counting parameter $e^{2\lambda}$ is the square g_c^2 of the physical coupling constant; see Eq. (1.13). On the other hand, we have the loop expansion of the 2d σ -model which lives on the string world-sheet: the σ -model loop-counting parameter is α' , as we saw when discussion the world-sheet β -functions. These σ -model perturbative contributions are called α' -corrections. From the point of view of spacetime physics, the α' -corrections are *classical higher derivative* corrections to the low-energy effective Lagrangian.

$z = 0$. In the previous example we considered graviton states, but it is obvious that the relation should hold for the infinite tower of physical states we constructed in the light-cone gauge. This correspondence holds because the world-sheet theory is conformal (superconformal in the superstring case), and in conformal field theory there is an isomorphism between states in the Hilbert space and local operators; see Sect. 2.2.2. The world-sheet theory of the string is in addition a gauge system, and the isomorphism will be between BRST-invariant physical states and BRST-invariant physical local operators (see Chap. 3). BRST-invariant local operators whose insertion creates on-shell physical states are called *vertices*.

To complete the construction of the (perturbative) string theory, we need to consider the BRST quantization and to construct vertices for all physical states, then learn how to compute their relevant correlations on surfaces of genus g , and finally how to integrate them over the moduli spaces $\mathcal{M}_{g,n}$ ($\mathcal{SM}_{g,n}$ for the superstring).

Here “perturbative” means order by order in the expansion of the physical amplitude in contributions from world-sheets of genus g , which represent the quantum correction at g string loops.

Since the world-sheet theory is conformal (superconformal), in order to define and compute all relevant quantities in an efficient way we need the powerful tools of 2d conformal field theory. This will be the subject of the next chapter.

Note 1.9 The string perturbative series itself is not expected to converge for a general process. That is, the series is only asymptotic for small string coupling, namely around backgrounds where the v.e.v. of the dilaton $\langle\Phi\rangle \rightarrow -\infty$. This fact may be equivalently stated as a *positive* assertion: ***superstring theory contains spectacular non-perturbative phenomena***. Some of them will be discussed at length in the final chapters of this textbook. These non-perturbative phenomena are needed to get a fully satisfactory quantum **Theory of Everything**.

References

1. Green, M.B., Schwarz, J., Witten, E.: Superstring Theory, vol. 1, 2. Cambridge University Press, Cambridge (1987)
2. Polchinski, J.: String Theory, vols. 1, 2. Cambridge University Press, Cambridge (2003)
3. Becker, K., Becker, M., Schwarz, J.: String Theory and M-Theory. Cambridge University Press, Cambridge (2007)
4. Blumenhagen, R., Lüst, D., Theisen, S.: Basic Concepts of String Theory. Springer (2013)
5. Polyakov, A.M.: Gauge Fields and Strings, vol. 3. Contemporary Concepts in Physics. Harwood (1987)
6. Schwarz, J.H.: Superstring theory. Phys. Rep. **89**, 223–322 (1982)
7. Zwiebach, B.: A First Course in String Theory. Cambridge University Press (2003)
8. Kiritsis, E.: String Theory in a Nutshell. Princeton University Press (2019)
9. Johnson, C.V.: D-Branes. Cambridge University Press (2003)
10. Uranga, A., Ibáñez, L.: String Theory and Particle Physics: An Introduction to String Phenomenology. Cambridge University Press (2012)

11. Di Vecchia, P.: The Birth of String Theory, vol. 737. Lectures Notes in Physics, pp. 59–118 (2008). [arXiv:0704.0101](#)
12. Di Vecchia, P., Schwimmer, A.: The Beginning of String Theory: A Historical Sketch, vol. 737. Lecture Notes Phys, pp. 119–136 (2008). [arXiv:0708.3940](#)
13. Schwarz, J.H.: The early history of string theory and supersymmetry. [arXiv:1201.0981](#)
14. 't Hooft, G.: A two-dimensional model for mesons. Nucl. Phys. **B75**, 461–470 (1974)
15. Weinberg, S., Witten, E.: Limits on massless particles. Phys. Lett. **B 96**, 59–62 (1980)
16. Weinberg, S.: Derivation of gauge invariance and the equivalence principle from Lorentz invariance of the S-matrix. Phys. Lett. **9**, 357–359 (1964)
17. Weinberg, S.: Photons and gravitons in S-matrix theory: derivation of charge conservation and equality of gravitational and inertial mass. Phys. Rev. **135**, B1049–B1056 (1964)
18. Weinberg, S.: Photons and gravitons in perturbation theory: derivation of Maxwell's and Einstein's equations. Phys. Rev. **138**, B98–B1002 (1965)
19. Polyakov, A.M.: Quantum geometry of bosonic strings. Phys. Lett. **B 103**, 207 (1981)
20. Alvarez, O.: Theory of strings with boundaries. Nucl. Phys. **B 216**, 125 (1983)
21. Friedan, D.: Introduction to Polyakov's string theory. In: Zuber, J.B., Stora, R. (eds.) Recent Advances in Field Theory and Statistical Mechanics. North Holland, Les Houches (1982)
22. Becchi, C., Rouet, A., Stora, R.: Renormalization of gauge theories. Ann. Phys. **98**, 287–321 (1976)
23. Tyutin, I.V.: Gauge Invariance in Field Theory and Statistical Physics in Operator Formalism. Lebedev Physics Institute preprint 39 (1975). [arXiv:0812.0580](#)
24. Kugo, T., Uehara, S.: General procedure of gauge fixing based on BRS invariance principle. Nucl. Phys. **B 197**, 378 (1982)
25. Becchi, C.: Introduction to BRS symmetry. [arXiv:hep-th/9607181](#)
26. Goddard, P., Goldstone, J., Rebbi, C., Thorn, C.B.: Quantum dynamics of a massless relativistic string. Nucl. Phys. **B 56**, 109 (1973)
27. Bjorken, J.D., Drell, S.D.: Relativistic Quantum Fields. McGraw-Hill (1965)
28. Postnikov, M.M.: Geometry VI. Riemannian Geometry, vol. 91. Encyclopaedia Mathematical Sciences. Springer (2001)
29. Lang, S.: Differential and Riemannian Manifolds. Springer (1995)
30. Hawking, S.W.: Zeta function regularization of path integrals in curved spacetime. Comm. Math. Phys. **55**, 133–148 (1977)
31. Berline, N., Getzler, E., Vergne, M.: Heat Kernels and Dirac Operators. Springer (2004)
32. NIST Digital Library of Mathematical Functions. <https://dlmf.nist.gov>
33. Göttsche, L.: The Betti numbers of the Hilbert scheme of points on a smooth projective surface. Math. Ann. **286**, 193–207 (1990)
34. Beauville, A.: Variétés Kähleriennes dont la première classe de Chern est nulle. J. Diff. Geom. **18**, 755–782 (1983)
35. Verbitsky, M.: Cohomology of compact hyperkaehler manifolds. [arXiv:alg-geom/9501001](#)
36. Polyakov, A.M.: Quantum geometry of fermionic strings. Phys. Lett. **B 103**, 211–213 (1981)
37. Cecotti, S.: Supersymmetric Field Theories. Geometric Structures and Dualities. Cambridge University Press (2015)
38. Griffiths, P., Harris, J.: Principles of Algebraic Geometry. Wiley (1978)
39. Hirsh, M.W.: Differential Topology. Springer (1991)
40. Prasolov, V.V.: Elements of Combinatorial and Differential Topology, vol. 74. Graduate Studies in Mathematic. American Mathematical Society (2006)
41. Yano, K.: The Theory of Lie Derivatives and Its Applications. Nabu Press (2011)
42. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry, vols. 1 and 2. Wiley (1963)
43. Chern, S.S.: Complex Mmanifolds without Potential Theory (with an Appendix in the Geometry of Characteristic Classes). Springer (1979)
44. Huybrechts, D.: Complex Geometry. An Introduction. Universitext, Springer (2005)
45. Kobayashi, S.: Differential Geometry of Complex Vector Bundles. Princeton University Press (1987)

46. Kobayashi, S.: Transformation Groups in Differential Geometry. Classics in Mathematics. Springer (1995)
47. Faddeev, L.D., Popov, V.: Feynman diagrams for the Yang-Mills field. Phys. Lett. B **25**, 29 (1967)
48. Pauli, W., Villars, F.: On the invariant regularization in relativistic quantum theory. Rev. Mod. Phys. **21**, 434–444 (1949)
49. Teschner, J.: Liouville theory revisited. Class. Quant. Grav. **18**, R153–R222 (2001). [arXiv:hep-th/0104158](https://arxiv.org/abs/hep-th/0104158)
50. Nakayama, Yu.: Liouville field theory: a decade after the revolution. Intern. J. Mod Phys. A **19**, 2771–2930 (2004). [arXiv:hep-th/0402009](https://arxiv.org/abs/hep-th/0402009)
51. Greene, R.E., Krantz, S.G.: Function Theory of One Complex Variable, vol. 40. Graduate Studies in Mathematics. AMS (2006)
52. Hartshorne, R.: Algebraic Geometry, vol. 52. Graduate Texts in Mathematics. Springer (1977)
53. Belavin, A.A., Polyakov, A.M., Schwartz, A.S., Tyupkin, Y.S.: Pseudoparticle solutions of the Yang-Mills equations. Phys. Lett. B **59**, 85–87 (1975)
54. 't Hooft, G.: Computation of the quantum effects due to a four-dimensional pseudoparticle. Phys. Rev. **D14**, 3432–3450 (1976)
55. Coleman, S.R.: The uses of instantons. Subnucl. Ser. **15**, 805 (1979)
56. Atiyah, M.F., Singer, I.M.: The index of elliptic operators on compact manifolds. Bull. Amer. Math. Soc. **69**, 422–433 (1963)
57. Atiyah, M.F.: Global theory of elliptic operators. In: Proceedings of International Conference on Functional Analysis and Related Topics (Tokyo, 1969), University of Tokyo (1970)
58. Adler, S.L., Bardeen, W.A.: Absence of higher order corrections in the anomalous axial vector divergence. Phys. Rev. **182**, 1517 (1969)
59. Farkas, H.M., Kra, I.: Riemann Surfaces, vol. 71. Graduate Texts in Mathematics. Springer (1992)
60. Hirzebruch, F.: Topological Methods in Algebraic Geometry. Classical in Mathematics. Springer (1995)
61. Harris, J., Morrison, I.: Moduli of Curves. Springer (1998)
62. Looijenga, E.: A minicourse on moduli of curves, proceedings of School of Algebraic Geometry (Trieste 1999), vol. 1. ITCF Lecture Notes, pp. 267–2091 (2000). <http://users.ictp.it>
63. Kodaira, K.: Complex Manifolds and Deformations of Complex Structures. Springer (1986)
64. Liu, K., Sun, X., Yang, X., Yau, S.-T.: Curvatures of moduli space of curves and applications. [arXiv:1312.6932](https://arxiv.org/abs/1312.6932)
65. Mirzakhani, M.: Simple geodesics and Weil-Persson volumes of moduli spaces of bordered Riemann surfaces. Inv. Math. **167**, 179–222 (2007)
66. Dijkgraaf, R., Witten, E.: Developments in topological gravity. [arXiv:1804.03275](https://arxiv.org/abs/1804.03275)
67. Mirzakhani, M.: Weil-Petersson volumes and intersection theory on the moduli space of curves. J. AMS **20**, 1–23 (2007)
68. Penner, R.: Weil-Petersson volumes. J. Differ. Geom. **35**, 559–608 (1992)
69. Witten, E.: On the structure of the topological phase of two-dimensional gravity. Nucl. Phys. B **340**, 281–332 (1990)
70. Witten, E.: Two-dimensional gravity and intersection theory on moduli space. Surv. Diff. Geom. **1**, 243–310 (1991)
71. Yau, S.T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampere equation. I. Commun. Pre Appl. Math. **31**, 339–411 (1978)
72. Helgason, S.: Differential Geometry, Lie Groups and Symmetric Spaces . Academic Press (1978)
73. Wess, J., Zumino, B.: Supergauge transformations in four dimensions. Nucl. Phys. B **70**, 39–50 (1974)
74. Green, M.B., Schwarz, J.H.: Covariant description of superstrings. Phys. Lett. B **136**, 367 (1984)
75. Berkovits, N., Gomez, H.: An Introduction to Pure Spinor Superstring Theory. [arXiv:1711.09966](https://arxiv.org/abs/1711.09966)

76. Brink, L., Di Vecchia, P., Howe, P.S.: A locally supersymmetric and reparametrization invariant action for the spinning string. *Phys. Lett. B* **65**, 471–474 (1976)
77. Brink, L., Schwarz, J.: Local complex supersymmetry in two dimensions. *Nucl. Phys. B* **121**, 285–295 (1977)
78. Brink, L., Deser, S., Zumino, B., Di Vecchia, P., Howe, P.S.: Local supersymmetry for spinning particles. *Phys. Lett. B* **64**, 435 (1976) [erratum: *Phys. Lett. B* **68**, 488 (1977)]
79. Atiyah, M.F.: Riemann surfaces and spin structures. *Annales scientifiques de l'É.N.S.* 4^{te} série, tome **4**(1), 47–62 (1971)
80. Seiberg, N., Witten, E.: Spin structures in string theory. *Nucl. Phys. B* **276**, 272 (1986)
81. Penrose, R., Rindler, W.: *Spinors and space-time*, vol. 2. Cambridge Monographs on Mathematical Physics. Cambridge University Press, Cambridge (1986)
82. H. Baum, Conformal Killing spinors and special geometric structures in Lorentzian geometry – A survey. In: *Proceedings of the Workshop on Special Geometric Structures in String Theory*, Bonn (2001). *Proceedings Archive of the EMS Electronic Library of Math*
83. Baum, H.: Conformal Killing spinors and the holonomy problem in Lorentzian geometry – a survey of new results. In: *Symmetries and Overdetermined Systems of Partial Differential Equations*, IMA Vol. Math. Appl., vol. 144. Springer, New York, pp. 251–264 (2008)
84. Leitner, F.: About twistor spinors with zero in Lorentzian geometry. *SIGMA* **5**, 079 (2009). [arXiv:math/0406298](https://arxiv.org/abs/math/0406298)
85. Kac, V.G.: Classification of simple Lie superalgebras, *Functional. Anal. Appl.* **9**, 263–265 (1975)
86. Kac, V.G.: Lie superalgebras. *Adv. Math.* **26**, 8 (1977)
87. Witten, E.: Notes on super-Riemann surfaces and their moduli. *Pure Appl. Math. Quart.* **15**(1), 57–211 (2019). [arXiv:1209.2459](https://arxiv.org/abs/1209.2459)
88. Witten, E.: Superstring perturbation theory revisited. [arXiv:1209.5461](https://arxiv.org/abs/1209.5461)
89. Witten, E.: Notes on supermanifolds and integration. *Pure Appl. Math. Quart.* **15**(1), 3–56 (2019). [arXiv:1209.2199](https://arxiv.org/abs/1209.2199)
90. Witten, E.: Superstring perturbation theory via super-Riemann surfaces: an overview. [arXiv:1304.2832](https://arxiv.org/abs/1304.2832)
91. Witten, E.: Notes on holomorphic string and superstring theory measures of low genus. [arXiv:1306.3621](https://arxiv.org/abs/1306.3621)
92. Callan, C.G., Friedan, D., Martinec, E.J., Perry, M.J.: Strings in background fields. *Nucl. Phys. B* **262**, 593 (1985)
93. Callan, C.G., Thorlacius, L.: Sigma models and string theory. In: Jevicki, A., Tan, C.-I. (eds.) *Particles, Strings and Supernovae*, TASI 1988, pp. 795–878, World-Scientific (1989)
94. Tseytlin, A.A.: Sigma model approach to string theory. *Int. J. Mod. Phys. A* **4**, 1257 (1989)
95. Jöst, J.: *Riemannian Geometry and Geometric Analysis*, 7th edn. Springer (2007)
96. Hélein, F., Wood, J.C.: Harmonic maps. In: Krupka, D., Saunders, D. (eds.) *Handbook of Global Analysis*, pp. 417–491. Elsevier science (2007)
97. Bott, R., Tu, L.W.: *Differential Forms in Algebraic Topology*, vol. 82. Graduate Texts in Mathematics. Springer (1982)
98. Berger, M., Gauduchon, P., Mazet, E.: *Le spectre d'une variété riemannienne*, vol. 194. *Lecture Notes in Mathematics*. Springer (1971)
99. Eisenhart, P.: *Riemannian Geometry*. Princeton University Press (1997)
100. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](https://arxiv.org/abs/1711.00864)
101. Weinberg, S.: *The Quantum Theory of Fields*, vol. 1, 2, 3. Cambridge University Press (1995)
102. Besse, A.: *Einstein Manifolds*. Springer (1987)
103. Hamada, Y., Vafa, C.: 8d supergravity, reconstruction of internal geometry and the swampland. *JHEP* **06**, 178 (2021). [arXiv:2104.05724](https://arxiv.org/abs/2104.05724)
104. DeWitt, B.S.: *Dynamical Theory of Groups and Fields*. Gordon and Breach (1965)
105. DeWitt, B.S.: Quantum theory of gravity II: the manifestly covariant theory. *Phys. Rev.* **162**, 1195 (1967)

Chapter 2

Review of 2d Conformal Field Theories



Abstract In this chapter we give an in-depth introduction to 2d conformal and superconformal field theories (CFT and SCFT) with special emphasis on ideas and techniques which are essential in string theory to define the theory, check its consistency, and compute concrete physical observables. Non-standard free systems, current algebras, and bosonization techniques are described in great detail. General references for this chapter are [1–10].

2.1 Spacetime Symmetries in QFT

We start from general background on spacetime symmetries in QFT. The reader may prefer to jump directly to Sect. 2.2.

Order of Transitivity

An important property of a spacetime symmetry group G is its *order of transitivity* $k \in \mathbb{N}$ [11]. A symmetry group G acting on the spacetime manifold M is *k-transitive* iff, given two arbitrary sets of k distinct points

$$\{x_1, \dots, x_k\} \quad \text{and} \quad \{y_1, \dots, y_k\}, \quad (2.1)$$

there is an element $\xi \in G$ such that $\xi(x_i) = y_i$ for all $i = 1, \dots, k$. If a QFT defined in the spacetime M is invariant under a k -transitive symmetry G , all its n -point functions with $n \leq k$

$$\langle \mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_n}(x_n) \rangle, \quad n \leq k \quad (2.2)$$

are determined by the symmetry up to a few numerical constants. Indeed, using the symmetry we may relate the function (2.2) to the correlator with the operators inserted at some pre-defined convenient points $\{\hat{x}_1, \dots, \hat{x}_n\}$, and the n -point function is fully determined by the numbers

$$\lambda_{i_1 i_2 \dots i_n} \equiv \langle \mathcal{O}_{i_1}(\hat{x}_1) \cdots \mathcal{O}_{i_n}(\hat{x}_n) \rangle. \quad (2.3)$$

A QFT is *almost-topological* if the order of transitivity of its spacetime symmetry G is infinite. In this case all correlations functions are essentially determined, and the QFT is explicitly solvable. An almost-topological QFT does not describe *local* degrees of freedom but may be highly non-trivial.

Exercise 2.1 Show that all *local* observables $\phi(x)$ of a quasi-topological QFT are scalar fields.

Spacetime Symmetries and the Energy–Momentum Tensor

Although most of the following arguments hold, *mutatis mutandis*, for all signatures of the spacetime manifold M , for definiteness we work in Euclidean signature. We take the spacetime to be a connected, geodesically complete,¹ oriented Riemannian manifold M with metric g_{ab} , and consider a QFT on the (Euclidean) spacetime M . As part of the definition of what we mean by “a QFT”, we require the existence² of a local energy–momentum tensor T_{ab} which is (covariantly³) *conserved* and *symmetric*

$$D^a T_{ab} = 0, \quad T_{ab} = T_{ba}, \quad (2.4)$$

together with an algebra $\mathfrak{A}$ of local (distribution-valued) quantum operators $\mathcal{O}_j(x)$ which transform as tensors under local diffeomorphisms.

Killing Vectors and Isometries Recall that a *vector field* $v \equiv v^a \partial_a$ (a smooth section of the tangent bundle $TM \rightarrow M$) is a *Killing vector* iff it satisfies the equation

$$\mathfrak{L}_v g_{ab} \equiv D_a v_b + D_b v_a = 0. \quad (2.5)$$

Let v be *any* vector field (not necessarily Killing); its *flow* is the solution to the differential equation

$$\frac{d}{dt} f_v^a(x, t) = v^b(x) \partial_b f_v^a(x, t), \quad \text{with initial condition } f_v^a(x, 0) = x^a, \quad (2.6)$$

where t is a real parameter and the x^a 's are local coordinates. The vector field v is *complete* iff the solution to the problem (2.6) exists *globally* on M for all $t \in \mathbb{R}$. A complete vector field v generates a *one-parameter family* of diffeomorphisms, $f_v(t): M \rightarrow M$ which we write as the exponential of the vector v

¹ On a Riemannian manifold there are three equivalent definitions of *completeness*:

theorem (Hopf–Rinow [12]). *For a Riemannian manifold M the following are equivalent:*

- (1) *M is geodesically complete, that is, each maximal geodesic $\gamma(t)$ is defined for all $t \in \mathbb{R}$;*
- (2) *M is complete as a metric space, that is, all Cauchy sequences converge;*
- (3) *the bounded subsets of M are relatively compact.*

If one (hence all) condition is satisfied, given two points $p, p' \in M$ there is at least one geodesic connecting them.

² We do not exclude the case where T_{ab} is the zero operator: this holds for a *topological field theory* (TFT) [13]. We are mostly interested in situations where $T_{ab} \neq 0$.

³ D_a is the covariant derivative of Riemannian geometry (the Levi-Civita connection).

$$f_v(t) \equiv \exp(tv), \quad f_v(t+s) = f_v(t) \circ f_v(s), \quad f_v(0) = \text{Id}_M. \quad (2.7)$$

On the contrary, the flow of a *non*-complete vector field produces a map which is *only locally defined* in some domain $D \subset M$.

If Y is any *tensor field* of type (k, l) on M (i.e. a smooth section of the vector bundle $TM^{\otimes k} \otimes T^*M^{\otimes l}$), we have

$$\frac{d}{dt} f_v(t)^* Y = \mathbb{L}_v f_v(t)^* Y \quad (2.8)$$

where $\mathbb{L}_v$ is the *Lie derivative* of the tensor field $f_v(t)^* Y$ along the vector field v [14].

When the vector field v is Killing, more is true.

Proposition 2.1 (see [11]) *If the Riemannian manifold M is complete, the Killing vectors v are complete and generate one-parameter families of global isometries $f_v(t): M \rightarrow M$ ($t \in \mathbb{R}$). The isometry group $\text{Iso}(M)$ of a Riemannian manifold is a Lie group. Its Lie algebra $\mathfrak{iso}(M)$ is generated by the Killing vectors (2.5). Moreover,*

$$\dim \mathfrak{iso}(M) \leq \frac{d(d+1)}{2} \quad \text{with equality for maximally symmetric spaces.} \quad (2.9)$$

Thus the Killing vectors generate (by exponentiation) the connected component $\text{Iso}(M)^0$ of the isometry group $\text{Iso}(M)$. By definition isometry groups are 1-transitive when the Riemannian manifold M is *homogeneous*, and 0-transitive otherwise.

Conserved Currents and Ward Identities The two properties (2.4) of the energy-momentum tensor imply that, for all *Killing vectors* $v \equiv v^a \partial_a$, the associated currents

$$T_a^v \stackrel{\text{def}}{=} T_{ab} v^b \quad (2.10)$$

are (covariantly) conserved: indeed

$$D^a (T_{ab} v^b) = (D^a T_{ab}) v^b + \frac{1}{2} T_{ab} (D^a v^b + D^b v^a) = 0. \quad (2.11)$$

We see the currents T_a^v as *closed* $(d-1)$ -forms

$$T^v \stackrel{\text{def}}{=} *(T_{ab} v^b dx^a), \quad dT^v = 0, \quad (2.12)$$

where $*$ is the Hodge-star defined by the orientation and the metric g_{ab} [12].

The equation $dT^v(x) = 0$ holds in the *distribution-operator* sense, that is, when $dT^v(x)$ is inserted in a correlation function with other (non-trivial⁴) local operators

$$\langle dT^v(x) \mathcal{O}_1(y_1) \cdots \mathcal{O}_s(y_s) \rangle_M, \quad (2.13)$$

⁴ By *non-trivial* local operator we mean a local operator which is not proportional to the identity.

the amplitude vanishes as long as $x \neq y_j$ for $j = 1, \dots, s$. However, as a distribution in spacetime, (2.13) is not necessarily zero: in general we may only conclude that **the amplitude (2.13) is a distribution in x with support on the finite set of points $\{y_1, \dots, y_s\}$** . The local contributions at the y_j 's are called *contact terms*. The precise contact terms depend on the quantization scheme (i.e. on the choices of local counterterms in the background fields and operator improvements).

A distribution-operator in x with support at y is a finite sum of the form⁵

$$\sum_{k=0}^{\ell} D^{a_1} \dots D^{a_k} [\phi_{a_1 \dots a_k}^k \delta^{(d)}(x - y)], \quad (2.14)$$

with operator coefficients $\phi_{a_1 \dots a_k}^k$ (cf. [15] Sect.2.3). We write the contact terms as

$$dT^v(x) O_j(y) = -\delta_v O_j(y) \delta^{(d)}(x - y) + \text{derivatives of } \delta^{(d)}(x - y), \quad (2.15)$$

which we see as the *definition* of the local operator $\delta_v O_j(y)$ called the “infinitesimal variation of $O_j(x)$ along the vector field v ”. In the absence of anomalies and for “good” local operators which transform as tensors under local reparametrizations, we have

$$\delta_v O_j(y) = \mathcal{L}_v O_j(y). \quad (2.16)$$

If $S \subset M$ is a codimension-1 oriented hypersurface with no boundary, $\partial S = \emptyset$, we define the “conserved” charge operator

$$Q^v(S) = \int_S T^v. \quad (2.17)$$

$Q^v(S)$ is actually conserved if S is compact or if S is non-compact with appropriate boundary conditions at infinity. Technically *conserved* means the following:

Let $\mathring{M} \equiv M \setminus \{y_1, \dots, y_s\} \subset M$ be the (open) complement in spacetime of the collection of points where we insert non-trivial local operators. Let $S \subset \mathring{M}$ be a compact oriented codimension-1 submanifold without boundary (i.e. $\partial S = \emptyset$). Then the operator $Q^v(S)$ depends on S only through its homology class $[S] \in H_{d-1}(\mathring{M}, \mathbb{Z})$ in $\mathring{M}$.

Indeed, if S and S' are homologous in $\mathring{M}$, i.e. $S' - S = \partial L$ for some d -chain $L \subset \mathring{M}$,

$$Q^v(S') - Q^v(S) = \int_{S'} T^v - \int_S T^v \equiv \int_{\partial L} T^v = \int_L dT^v = 0, \quad (2.18)$$

by Stokes theorem [16]. We stress that L should be a cycle in $\mathring{M}$ rather than in M ; otherwise contact terms from operator insertions in L contribute to the RHS of (2.18).

⁵ Here and below $\delta^{(d)}(x - y)$ is a shorthand for $\delta(x^1 - y^1) \delta(x^2 - y^2) \dots \delta(x^d - y^d)$ where x^a, y^a ($a = 1, \dots, d$) are the coordinates of two points in M .

Let S be a closed oriented $(d-1)$ -cycle whose homology class $[S]$ is trivial in M . Then S splits the spacetime M in an interior part I (with $\partial I = S$) and an exterior part E (with $\partial E = -S \cup \partial M$) such that $M = I \cup E$ and $S = I \cap E$. We focus on correlation functions of the form

$$\left\langle \mathcal{Q}^v(S) \mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_s}(x_s) \right\rangle_M \equiv \left\langle \int_S T^v \mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_s}(x_s) \right\rangle_M, \quad (2.19)$$

where the points $x_j \in I$ for $j = 1, \dots, \ell$, while $x_j \in E$ for $j = \ell + 1, \dots, s$. Then S is homologous in $\mathring{M}$ to minus the sum of the boundaries of small spheres centered at the points x_j in the interior I . Hence, for each Killing vector v and closed oriented hypersurface $S \subset \mathring{M}$, we get an associated *Ward identity*

$$\left\langle \mathcal{Q}^v(S) \mathcal{O}_{i_1}(x_1) \cdots \mathcal{O}_{i_s}(x_s) \right\rangle_M = \sum_{j=1}^{\ell} \left\langle \mathcal{O}_{i_1}(x_1) \cdots \delta_v \mathcal{O}_{i_j}(x_j) \cdots \mathcal{O}_{i_s}(x_s) \right\rangle_M. \quad (2.20)$$

Energy–Momentum Tensors with Special Properties

If the energy–momentum tensor T_{ab} enjoys additional algebraic properties besides Eq. (2.4), we have more conserved currents of the form (2.10).

T_{ab} decomposes in two $O(d)$ representations, the trace and the traceless part, which are absolutely irreducible for $d > 2$. We assume that the background spacetime geometry consists only of the Riemannian metric g_{ab} and (possibly) the orientation form ε . We rule out all other background tensor fields which would reduce the structure group⁶ of spacetime to a *proper* subgroup $G \subset SO(d)$. In this standard situation the non-zero components of T_{ab} should form complete $SO(d)$ -representations. Thus (a part for the TFT case where $T_{ab} \equiv 0$) we have two possibilities (in $d = 2$ the story is slightly richer):

- (1) T_{ab} transforms in the trivial $O(d)$ representation: $T_{ab} = g_{ab} T$ for a scalar T ;
- (2) T_{ab} contains only the symmetric traceless part, i.e. its trace vanishes $T^a_a = 0$.

The operator equalities $T_{ab} - g_{ab} T = 0$ (resp. $T^a_a = 0$) are always meant in the distribution-operator sense, i.e. modulo contact terms.

In situation (1) the current $T_{ab} v^b$ is conserved for all vectors v^b such that

$$0 = g^{ab} \mathfrak{L}_v g_{ab} \equiv v^a \partial_a \log \det g. \quad (2.21)$$

These vectors generate the *volume-preserving diffeomorphisms*. Thus in case (1) our QFT is invariant under all volume-preserving diffeomorphisms. Then, by a theorem of Moser [11], the partition function on a compact Riemannian space M depends only on the smooth structure of M and the number $\text{vol}(M)$. This spacetime group is ∞ -transitive [11], so the spacetime dependence of *all* n -point functions is determined by the symmetry and the QFT is exactly solvable: it is an almost-topological theory with no local dynamics. This is obvious from the equation

⁶ For G -structures on manifolds, see [11].

$$0 = D^a T_{ab} = g_{ab} \partial^a T, \quad (2.22)$$

which says that T is constant in spacetime so localized fluctuations cannot exist.

Exercise 2.2 Show that 2d Yang–Mills is a *non-trivial* QFT-invariant under all volume-preserving diffeomorphisms.

In this chapter we are interested in the opposite situation where only the traceless part of T_{ab} is non-zero.

2.2 Conformal Field Theory (CFT)

The opposite case is when T_{ab} does not contain the trivial representation, i.e. it is traceless, $T^a_a = 0$. In this situation the currents T_a^v in Eq. (2.10) are conserved for v^a a *conformal Killing vector* (CKV), that is, a vector field $v \equiv v^a \partial_a$ which satisfies the CKV differential equation

$$\mathfrak{L}_v g_{ab} \equiv D_a v_b + D_b v_a = 2\lambda(x) g_{ab} \quad \text{for some function } \lambda(x). \quad (2.23)$$

Indeed,

$$D^a (T_{ab} v^b) = T_{ab} D^a v^b = \frac{1}{2} T_{ab} (D^a v^b + D^b v^a) = \lambda T_{ab} g^{ab} = 0. \quad (2.24)$$

In general a CKV generates (by exponentiation) only a *local* conformal transformation of M . When the CKV v is also *complete* in the sense of Sect. 2.1, it generates a *globally defined* one-parameter group of conformal maps $f_v(t): M \rightarrow M$ ($t \in \mathbb{R}$).

The charges $Q^v(S)$ associated with *complete* CKVs v ,

$$Q^v(S) = \int_S T^v, \quad (2.25)$$

generate a Lie group which is isomorphic to the connected component $\text{Conf}(M)^0$ of the group $\text{Conf}(M)$ of conformal automorphisms of M . In other words, *all* global conformal automorphisms of M (continuously connected to the identity) are space-time symmetries for a QFT iff the equation

$$T^a_a = 0 \quad (2.26)$$

holds as an operator statement. The QFTs which satisfy the condition (2.26) are called *conformal field theories* (CFT).

While *bona fide* symmetries in $\text{Conf}(M)^0$ are generated by *complete* CKVs, Eq. (2.24) says that the currents T_a^v are conserved for *all* CKV v . In conclusion: **non-complete CKVs do not generate symmetries, but they do lead to conserved**

currents T_a^v , hence to Ward identities for the quantum amplitudes. We stress that this statement rests on the assumption that the expression T^v in (2.12) may be defined at the quantum level as a *global* $(d - 1)$ -form.

2.2.1 Conformal Automorphisms and Equivalences

By definition, the *complete* CKVs generate one-parameter groups of *conformal automorphisms* of M , that is, *bona fide* symmetries of the conformal field theory. We recall the basic math facts.

Theorem 2.1 *Let M be a d -dimensional Riemannian manifold. The group $\text{Conf}(M)$ of conformal automorphisms of M is a Lie group when (a) $d \geq 3$, or when (b) $d = 2$ and M is compact. In these two situations one has*

$$\dim \text{Conf}(M) \leq \frac{(d+1)(d+2)}{2} \quad \begin{array}{l} \text{necessary condition for equality:} \\ M \text{ conformally flat} \\ \text{sufficient if } M \text{ is simply connected.} \end{array} \quad (2.27)$$

Moreover in these two cases: **(1)** the Lie algebra $\mathfrak{conf}(M)$ of $\text{Conf}(M)$ consists precisely of the CKV, and **(2)** when $\pi_1(M) \neq 0$ all CKVs are Killing vectors.

Proof Case (a) is **Theorem IV.6.1** of [11]. For case (b) we may assume with no loss that M is compact oriented; then M is a compact Kähler manifold and $\text{Conf}(M) \equiv \text{Aut}(M)$ the group of complex automorphisms. Then **Theorem III.1.1** of [11] yields all statements but the last one which is **Theorem IV.7.5** of [11] since each conformal class contains a metric with constant scalar curvature because the Yamabe problem has a positive answer [17, 18]. $\square$

For instance, if $M = S^d$ the group of conformal automorphisms is

$$\text{Conf}(S^d) = SO(d+1, 1), \quad (2.28)$$

and the inequality in (2.27) is saturated. In fact, saturation of the inequality implies $M = S^d$ when $d \geq 3$ or M is compact. The simplest way to see (2.28) is to write S^d as a real quadric in $\mathbb{RP}^{d+1}$

$$X_0^2 - \sum_{i=1}^{d+1} X_i^2 = 0 \quad (2.29)$$

endowed with its canonical metric induced from the ambient metric. The projective action of $SO(d+1, 1)$ is clearly conformal. In **BOX 2.1** it is proven that the order of transitivity of $SO(d+1, 1)$ acting on S^d is 3, so **in a (Euclidean) CFT on the d -dimensional sphere S^d , all $n \leq 3$ point functions are determined by symmetry up to a few numerical constants.**

The Lucky Case of the String World-Sheet

From **Theorem 2.1** we see that there is one exceptional case: (Euclidean) CFT on a *non-compact* Riemannian 2-manifold Σ which we may assume *oriented*—hence *complex*—by replacing it, if necessary, with its orientable double $\check{\Sigma}$.

This exception is *quite spectacular*: the space of conformal Killing vectors on a non-compact complex manifold Σ of (complex) dimension 1 is always *infinite-dimensional*.⁷ Indeed, we can fix any finite⁸ number of points $p_i \in \Sigma$ and find a holomorphic vector $v(z)\partial_z$ on Σ with prescribed values of the first n_i derivatives at each point p_i for any given collection of positive integers $\{n_i\}$.

BOX 2.1 - More on the Euclidean conformal group $SO(d+1, 1)$

We change variables and write the quadric (2.29) in real projective space as

$$2X_0X_{d+1} - \sum_{i=1}^d X_i^2 = 0. \quad (*)$$

By the Iwasawa decomposition theorem, all elements of the group $SO(d+1, 1)$ may be written in a *unique* way as a product

$$S \cdot \begin{pmatrix} \lambda & \lambda w^t & \lambda w^t w/2 \\ 0 & \mathbf{1}_d & w \\ 0 & 0 & \lambda^{-1} \end{pmatrix} \quad S \in SO(d+1), \lambda \in \mathbb{R}_{>0}, w \in \mathbb{R}^d. \quad \ddagger$$

S is an isometry of the sphere S^d , so the group of conformal automorphisms of S^d modulo isometries is the group T of triangular matrices displayed in the equation ($\ddagger$). By a $SO(d+1)$ rotation we may map any point $x_1 \in S^d$ to the North pole n of homogeneous coordinates $X_i = X_{d+1} = 0, X_0 \neq 0$. The group T fixes the North pole. Let $x_2 \neq x_1$ be a second point of homogeneous coordinates (X_0, X_i, X_{d+1}) ; we must have $X_{d+1} \neq 0$ since otherwise $X_i = 0$ by (*) and $x_2 \equiv x_1$. The element of T with $\lambda = 1$ and $w_i = -X_i/X_{d+1}$ maps x_2 to the South pole s of homogeneous coordinates $X_0 = X_i = 0, X_{d+1} \neq 0$. The two poles are preserved by the subgroup $SO(d) \times \mathbb{R}_{>0}$ which acts simply transitively on the points $x_3 \neq n, s$. We conclude that the order of transitivity of $SO(d+1, 1)$ is 3.

While almost all vectors $v(z)\partial_z$ in this “huge” infinite-dimensional space *do not* generate global symmetries,⁹ they *do lead* to conserved currents for all CFTs quantized on Σ and hence to valid Ward identities for the physical amplitudes on the non-compact world-sheet Σ .

As we saw in the last section of Chap. 1, in string theory all g -loop contributions to a non-vacuum physical amplitude are described by some CFT on a non-compact world-sheet Σ : all such amplitudes satisfy an *infinite family* of Ward identities which are quite helpful in computing them. *String theory is quite smart*: it exploits all possible exceptions to make itself physically consistent and, in a sense, “simple”.

⁷ Σ is *Stein* (see, for example, p. 134 of [19]) and hence the dimension of the space of holomorphic vector fields is infinite-dimensional by Cartan Theorem A (see [19, 20]).

⁸ Even an infinite number of them, provided they have no accumulation point in Σ .

⁹ The global symmetries form a finite-dimensional Lie group, whereas the CKV space is infinite-dimensional for Σ non-compact.

Conformal Equivalences

Up to now we considered conformal *symmetries*, i.e. automorphisms of the conformal structure of M which relate two observables of the CFT quantized in M . More generally, we are interested in conformal *equivalences* which relate an observable of the CFT on the Riemannian manifold¹⁰ (M, g) to a corresponding observable of the CFT quantized on a *different* Riemannian manifold (N, h) which is *conformally equivalent* to (M, g) , i.e. related by a diffeomorphism $f: M \rightarrow N$ such that

$$g(x)_{\alpha\beta} = e^{2\phi(x)} h(f(x))_{ij} \partial_{x^\alpha} f(x)^i \partial_{x^\beta} f(x)^j \quad (2.30)$$

for some function $\phi: M \rightarrow \mathbb{R}$. M and N are the same abstract manifold with two different Riemannian structures whose underlying conformal structures are equal. Since the equation $T^a_a = 0$ entails that the QFT is invariant under deformations of the metric which preserve its conformal class, the quantizations of the CFT on M and N produce physically equivalent systems with the same value for all observables which, however, now have different “geometric” interpretation in the two manifolds.

Example: $\mathbb{R}^d$ versus S^d

Flat Euclidean space $\mathbb{R}^d$ is conformally equivalent to the *punctured* d -sphere $S^d \setminus \{\text{South pole}\}$. Indeed, the “round” metric $d\Omega_d^2$ on the unit d -sphere S^d is defined recursively in d by

$$d\Omega_d^2 = d\theta^2 + \sin^2 \theta d\Omega_{d-1}^2, \quad 0 \leq \theta \leq \pi. \quad (2.31)$$

Writing $r = \tan(\theta/2)$ we have

$$d\Omega_d^2 = 4 \cos^4(\theta/2) (dr^2 + r^2 d\Omega_{d-1}^2), \quad (2.32)$$

which, up to the overall factor $4 \cos^4(\theta/2)$, is the flat metric on $\mathbb{R}^d$ written in polar coordinates. The South pole point $\theta = \pi$ is pushed to infinite distance in $\mathbb{R}^d$. We also say that S^d is the *conformal compactification* of $\mathbb{R}^d$ obtained by adding to $\mathbb{R}^d$ the “point at infinity” $r = \infty$.

Combining this observation with the results in **BOX 2.1**, we get the following: ***in an Euclidean conformal field theory (CFT) the functional form of all $n \leq 3$ point functions on $\mathbb{R}^d$ is determined by the symmetry.***

¹⁰ A *Riemannian manifold* is a pair (M, g) where M is a manifold and g is a (positive-definite) metric on the tangent bundle TM . Most of the time we shall be less pedantic and write just M .

2.2.2 Radial Quantization and the State-Operator Isomorphism

Consider a d -dimensional CFT quantized on the d -dimensional infinite cylinder

$$\text{Cy}_d \equiv \mathbb{R} \times S^{d-1}, \quad (2.33)$$

where the coordinate t on the factor $\mathbb{R}$ is identified with the Euclidean time. The standard (i.e. symmetric¹¹) metric on Cy_d is

$$ds^2 = dt^2 + d\Omega_{d-1}^2. \quad (2.34)$$

Changing coordinate to $r = e^t$ the metric becomes

$$ds^2 = \frac{1}{r^2} (dr^2 + r^2 d\Omega_{d-1}^2) \quad (2.35)$$

which, up to the overall factor r^{-2} , is again the flat metric on $\mathbb{R}^d$ written in polar coordinates. Thus, a CFT quantized on the cylinder Cy_d (with Euclidean time t) is equivalent to the same QFT quantized on $\mathbb{R}^d$ where now we take as Euclidean time $t \equiv \log r$, the logarithm of the radial coordinate r . This setup is called *radial quantization* [21]: it is the most convenient viewpoint when dealing with a CFT.

The Hamiltonian on the cylinder, H_{cyl} , acts on the Hilbert space $\mathbf{H}_{S^{d-1}}$ of states defined on a constant-time slice $\{t\} \times S^{d-1}$ by generating an infinitesimal translation in time $t \mapsto t + \delta t$. H_{cyl} is mapped in $\mathbb{R}^d$ to the radial Hamiltonian H_{rad} which acts on the Cartesian coordinates x^i of $\mathbb{R}^d$ by the overall rescaling

$$x^i \mapsto e^{\delta t} x^i, \quad (2.36)$$

so that H_{rad} is the *scaling operator* which generates *dilatations*: more precisely

$$[H_{\text{cyl}}, \phi] = \Delta(\phi) \phi, \quad (2.37)$$

where $\Delta(\phi)$ is the scaling dimension of the local operator ϕ . Therefore H_{cyl} coincides with the scaling operator only up to an additive constant.

Radial quantization is much more regular than the usual QFT quantization in $\mathbb{R}^d$ where we take as constant-time slices hyperplanes of the form $\{t\} \times \mathbb{R}^{d-1}$. Its better properties follow from the fact that the fixed-time spatial slices $\{r\} \times S^{d-1}$ are now *compact*. In the usual quantization in $\mathbb{R}^d$ the Hamiltonian H_{tran} generates linear translations of $\mathbb{R}^d$: the spectrum of H_{tran} is then continuous, and its eigenstates are

¹¹ A metric is (locally) *symmetric* iff its Riemann tensor is covariantly constant: $D_i R_{jklm} = 0$.

non-normalizable. On the contrary, in the absence of “pathologies”,¹² the spectrum of H_{rad} is purely *discrete* and *bounded below*—in fact, for all $\beta > 0$

$$\exp[-\beta H_{\text{rad}}] \quad (2.38)$$

is a positive *compact* operator¹³ when acting on the Hilbert space $\mathbf{H}_{S^{d-1}}$ —hence its eigenvalues are discrete with zero as their only accumulation point, its eigenstates are *normalizable*, and its eigenspaces are *finite-dimensional*.

The asymptotic infinite past $t \rightarrow -\infty$ on the cylinder $\mathbf{Cy}_d$ is mapped to the origin $x^i = 0$ in $\mathbb{R}^d$, while the hypersurfaces of constant time are spheres centered at the origin. The asymptotic future $t \rightarrow +\infty$ is mapped to the point ∞ at infinity in the conformal compactification S^d of $\mathbb{R}^d$.

Consider a physical process on the cylinder $\mathbf{Cy}_d$. We have an initial state $|\text{in}\rangle$ at $t \rightarrow -\infty$ and a final state $\langle \text{out}|$ at $t \rightarrow +\infty$ which specify the **in/out** boundary conditions for the path integral which computes the physical amplitude $\langle \text{out}|\text{in}\rangle$. From the point of view of radial quantization in $\mathbb{R}^d$, the initial boundary condition is localized at the origin $r = 0$: more or less by definition, this boundary condition is equivalent to the insertion of a *local operator* $O_{\text{in}}(x^i)$ at the origin $x^i = 0$. This yields a map between the states of the theory quantized on the equal-time spatial hypersurface $\{r\} \times S^{d-1}$ and the local operators. This map is in fact a *linear isomorphism* between the Hilbert space $\mathbf{H}_{S^{d-1}}$ and the algebra $\mathfrak{A}$ of local operators

$$\mathbf{H}_{S^{d-1}} \simeq \mathfrak{A}, \quad |\text{in}\rangle \mapsto O_{\text{in}}. \quad (2.39)$$

An explicit way to realize this isomorphism is to consider the CFT path integral on the unit ball $B \subset \mathbb{R}^d$ with the local operator $O_{\text{in}}(x)$ inserted at the origin and the Dirichlet boundary conditions on the unit sphere $S^{d-1} \equiv \partial B$, i.e.

$$\Phi \Big|_{S^{d-1}} = \Phi_0 \quad \text{Here } \Phi \text{ is a short-hand to represent all fields we are integrating over in the path integral} \quad (2.40)$$

with

$$\Phi_0: S^{d-1} \rightarrow (\text{field space}) \quad (2.41)$$

an arbitrary field configuration on the radial equal-time surface S^{d-1} . The path integral with this insertion and b.c. produces a wave-functional $\Psi_{O_{\text{in}}}[\Phi_0]$ of the field configuration Φ_0 which is the Schrödinger representation of a state $|\text{in}\rangle \in \mathbf{H}_{S^{d-1}}$

¹² That is, when the CFT is unitary and non-degenerate. In practice this holds for *compact* CFTs. For an example of non-compact CFT with continuous spectrum of H_{rad} , see Sect. 2.4.

¹³ The operator is even *Hilbert–Schmidt*, i.e. $\text{Tr}[(e^{-\beta H_{\text{rad}}})^\dagger (e^{-\beta H_{\text{rad}}})] = \text{Tr}[e^{-2\beta H_{\text{rad}}}] < \infty$ a condition which says that the “radial canonical ensemble” (that is, the canonical ensemble in the finite-volume space S^{d-1}) is well-defined. The ordinary canonical ensemble is typically *not* well-defined unless we take the volume to be finite.

$$\langle \Phi_0 | O_{\text{in}} \rangle = \Phi_{O_{\text{in}}}[\Phi_0] \equiv \int_{\Phi|_{\partial B} = \Phi_0} [d\Phi] e^{-S[\Phi]} O_{\text{in}}(0). \quad (2.42)$$

This (functional) Schrödinger representation defines the map

$$\iota: \mathfrak{A} \rightarrow \mathbf{H}_{S^{d-1}}, \quad \iota: O_{\text{in}} \mapsto |O_{\text{in}}\rangle. \quad (2.43)$$

The identity operator 1 is mapped to the *vacuum* $|0\rangle$ whose wave-functional $\langle \Phi_0 | 0 \rangle$ is given by the path integral on B without insertions

$$\langle \Phi_0 | 0 \rangle = \int_{\Phi|_{\partial B} = \Phi_0} [d\Phi] e^{-S[\Phi]}. \quad (2.44)$$

Exercise 2.3 Show that $|0\rangle$ is invariant under the full conformal group $SO(d+1, 1)$.

The **Exercise** implies that the radial-quantization vacuum $|0\rangle$ is invariant by translations in $\mathbb{R}^d$, hence is the *same vacuum* as in ordinary quantization.

Inserting in the path integral (2.44) the local operator $O(x)$ with $x \in B$, we see that for an arbitrary local operator $O(x)$,

$$\iota: O(x) \mapsto |O\rangle \stackrel{\text{def}}{=} \lim_{x \rightarrow 0} O(x)|0\rangle. \quad (2.45)$$

We conclude

Fact 2.1 ι is an isomorphism called the (*conformal*) *state-operator correspondence*.

The state-operator correspondence is a basic tool in CFT and string theory.

The Dual State-Operator Correspondence The correspondence ι was deduced from the fact that the backward time evolution on the cylinder Cy_d becomes as $t \rightarrow -\infty$ the retraction to the origin in $\mathbb{R}^d$. Dually, forward time evolution on the cylinder becomes retraction to the point at infinity $\infty \in S^d$. The sphere has a symmetry which interchanges the two poles $0 \leftrightarrow \infty$, so the previous arguments apply to the $t \rightarrow +\infty$ limit, giving a dual operator-state correspondence for the *bra* vector space

$$O(z) \mapsto \langle O| = \lim_{y \rightarrow \infty} \langle 0|O(y) \quad (2.46)$$

where y is the appropriate local coordinate around ∞ on S^d .

Note 2.1 To avoid misunderstandings we stress that (2.46) is a $\mathbb{C}$ -linear isomorphism, that is, the *bra* $\langle O|$ depends linearly on O **not** anti-linearly as in the usual convention of QM. In other words, in the notation which is standard in CFT, the Hermitian conjugate of the *ket* state $|O\rangle$ is the *bra* state $\langle O^\dagger|$ which corresponds to the Hermitian conjugate operator $O(x)^\dagger$ of the local operator $O(x)$. With this CFT

convention the isomorphism ι intertwines Hermitian conjugation on states and operators. The Hilbert space inner product of the state $|\mathcal{O}_1\rangle$ with $|\mathcal{O}_2\rangle$ is the Hermitian form

$$\langle \mathcal{O}_1^\dagger | \mathcal{O}_2 \rangle = \langle \mathcal{O}_1^\dagger(\infty) \mathcal{O}_2(0) \rangle_{S^d}, \quad (2.47)$$

anti-linear in the first argument, linear in the second one.

2.2.3 Operator Product Expansions (OPE)

Let $\{\phi_\alpha\}_{\alpha \in A}$ be a topological basis of the local operator algebra $\mathfrak{A}$ whose elements ϕ_α have definite scaling dimension Δ_α . It may be convenient to choose the basis to be orthonormal with respect to the Hilbert space Hermitian product (2.47)

$$\langle \phi_\alpha^\dagger | \phi_\beta \rangle = \delta_{\alpha\beta}. \quad (2.48)$$

Consider two local operators ϕ_α and ϕ_β . We insert ϕ_β at the origin and ϕ_α in some point x with $0 < |x| < 1$. Performing the path integral over the unit ball B with the Dirichlet b.c. (2.40) we produce the Schrödinger representation of the state

$$\phi_\alpha(x) \phi_\beta(0) |0\rangle \in \mathbf{H}_{S^{d-1}}. \quad (2.49)$$

By the isomorphism ι , this state can also be produced by a single insertion of some local operator $\mathcal{O}(0)_{\alpha,\beta;x} \in \mathfrak{A}$ at the origin. The expansion of $\mathcal{O}(0)_{\alpha,\beta;x}$ in the basis $\{\phi_\alpha\}_{\alpha \in A}$ has the general form

$$\mathcal{O}(0)_{\alpha,\beta;x} = \sum_{\gamma \in A} f(x)_{\alpha\beta}^\gamma \phi_\gamma(0) \quad (2.50)$$

for some coefficient functions $f(x)_{\alpha\beta}^\gamma$ which depend on the operators ϕ_α, ϕ_β and the insertion point x of ϕ_α . The isomorphism ι then yields the operator equality

$$\phi_\alpha(x) \phi_\beta(0) = \sum_{\gamma \in A} f(x)_{\alpha\beta}^\gamma \phi_\gamma(0) \quad (2.51)$$

where the product in the LHS is multiplication in the associative operator algebra $\mathfrak{A}$. Since the CFT is translation-invariant in $\mathbb{R}^d$, we get the operator identity

$$\phi_\alpha(x) \phi_\beta(y) = \sum_{\gamma \in A} f(x-y)_{\alpha\beta}^\gamma \phi_\gamma(y) \quad (2.52)$$

called the *operator product expansion* (OPE) of the local operators $\phi_\alpha(x)$ and $\phi_\beta(y)$. The OPE is a basic tool in CFT (and also in QFT in general). We stress that Eq. (2.52)

is an *exact* formula, not an asymptotic series. In particular the sum in the RHS is *convergent* for $x \neq y$ (in the weak topology for $\mathfrak{A}$).

Equating the adjoint actions of H_{rad} on the two sides of Eq.(2.52), we get the scaling property of the coefficient functions

$$f(\lambda x)_{\alpha\beta}^\gamma = \lambda^{\Delta_\gamma - \Delta_\alpha - \Delta_\beta} f(x)_{\alpha\beta}^\gamma, \quad \text{for } \lambda \in \mathbb{R}_{>0}. \quad (2.53)$$

In a non-degenerate unitary CFT the dimension spectrum is discrete and bounded below, while the eigenspaces of H_{rad} are finite-dimensional. Therefore for given ϕ_α , ϕ_β there are only *finitely many* operators ϕ_γ with

$$\Delta_\gamma < \Delta_\alpha + \Delta_\beta \quad (2.54)$$

for which the function $f(x)_{\alpha\beta}^\gamma$ *blows up* as $x \rightarrow 0$. The corresponding terms in (2.52) are called the *singular part* of the OPE, while the other terms (which remain bounded as $x \rightarrow y$) are called its *regular part*. It is customary to write explicitly only the singular part of the OPE

$$\phi_\alpha(x) \phi_\beta(y) = \overbrace{\sum_{\Delta_\gamma < \Delta_\alpha + \Delta_\beta} f(x-y)_{\alpha\beta}^\gamma \phi(y)_\gamma}^{\text{finite sum}} + \text{regular}. \quad (2.55)$$

We also write $\sim$ to mean equality up to regular terms. The conformal field theory is (at least in principle) completely determined by the singular part of the OPEs.

The coefficient functions $f(x)_{\alpha\beta}^\gamma$ are not arbitrary. They are restricted by

- (1) the algebraic fact that the OPE product is *associative*;
- (2) the transformation properties under $SO(d+1, 1)$ of the fields ϕ_α ;
- (3) the requirement of unitarity.

Conversely, any set of functions $\{f(x)_{\alpha\beta}^\gamma\}_{\alpha,\beta,\gamma \in A}$ consistent with conditions (1)-(3) defines a CFT. This observation leads to a powerful approach to conformal field theory called the *conformal bootstrap*: for reviews, see [22–25].

Using the normalization (2.48) we get

$$\langle \phi_\gamma(\infty)^\dagger \phi_\alpha(x) \phi_\beta(0) \rangle_{S^d} = \langle \phi_\gamma^\dagger | \phi_\alpha(x) | \phi_\beta \rangle = f(x)_{\alpha\beta}^\gamma, \quad (2.56)$$

so the coefficient functions of the OPE are essentially the three-point functions on the sphere. Their dependence on x is then determined by the conformal transformations of the three operators ϕ_α , ϕ_β , ϕ_γ up to a few numerical constants.

2.3 CFT in 2d

In two dimensions conformal symmetry becomes both simpler and more powerful. General references for this section are [1, 3, 4, 7]. We assume the (connected) 2-manifold Σ to be *oriented* (otherwise we replace Σ with its oriented double $\check{\Sigma}$). Then Σ is a complex manifold of dimension 1 (a.k.a. 1-fold), and we use the local complex coordinate z . As shown in **BOX 1.4** the metric is automatically Kähler, $ds^2 = g_{z\bar{z}} dz d\bar{z}$, hence locally conformal to $dz d\bar{z}$.

The condition of zero trace becomes $T_{z\bar{z}} = 0$. The surviving components of the energy–momentum tensor are $T \equiv T_{zz}$ and $\bar{T} \equiv T_{\bar{z}\bar{z}}$. The conservation law now reads

$$0 = \partial_{\bar{z}} T_{zz} + \partial_z T_{\bar{z}\bar{z}} \equiv \partial_{\bar{z}} T, \quad (2.57)$$

so $T(z)$ is holomorphic ($\equiv$ left-moving) and $\bar{T}(\bar{z})$ anti-holomorphic ($\equiv$ right-moving). Purely left- or right-moving fields are called *chiral currents*, so the energy–momentum tensor of a CFT is a pair of left/right chiral currents.

Let $v(z)$ be a holomorphic vector field¹⁴ on Σ . The chiral current

$$J^v(z) \equiv \frac{1}{2\pi i} T(z) v(z) \quad (2.58)$$

is automatically conserved

$$\bar{\partial}(J^v(z)) \equiv \frac{1}{2\pi i} \bar{\partial}(T(z) v(z)) = 0, \quad (2.59)$$

in the distribution-operator sense (i.e. up to contact terms). Classically $J^v(z)$ transforms as a differential $(1, 0)$ -form under local holomorphic reparametrizations $z \rightarrow z'(z)$, that is,

$$J^v(z') dz' = J^v(z) dz \quad (\text{classically}). \quad (2.60)$$

However the quantum story is a bit subtler because of contact terms with the energy–momentum tensor. The contact terms will be computed momentarily. The bottom line is that $T(z)$ does not transform as a quadratic differential ($\equiv$ a holomorphic section of K^2) under *arbitrary* local holomorphic reparametrizations but only under a finite-dimensional space of “good” reparametrizations $z \mapsto z'(z)$ whose Jacobian dz'/dz satisfies a certain differential condition (see Sect. 2.3.3). Correspondingly, at the quantum level the chiral current $J^v(z)$ transforms as a $(1, 0)$ -form only for “good” local conformal motions. *A priori* one may worry that this quantum effect spoils the formal manipulations with the “charges”

$$Q^v(C) \equiv \oint_C J^v(z) \quad (2.61)$$

¹⁴ We suppress the holomorphic tangent space indices: $v(z)$ means $v(z)^z$.

we performed in Sect. 2.1, so that the Ward identities may be no longer valid. Luckily, this is not the case. Clearly, if Σ is covered by a single coordinate patch, the required manipulations of the contour C are valid by the Cauchy residue theorem in $\mathbb{C}$. Then our manipulations will be still legitimate if we use a “good” complex atlas, i.e. an open cover $\Sigma = \cup_a U_a$ by coordinates patches U_a , with local holomorphic coordinates z_a , such that whenever $U_a \cap U_b \neq \emptyset$ the function dz_a/dz_b satisfies the afore-mentioned differential condition. “Good” atlases exist for all Σ . For S^2 the standard two-patch atlas with $z_2 = 1/z_1$ is already good. For $\Sigma \neq S^2$ write $\Sigma = \tilde{\Sigma}/G$ where $\tilde{\Sigma}$ is the universal cover and $G \cong \pi_1(\Sigma)$ a discrete group. $\tilde{\Sigma}$ is either $\mathbb{C}$ or $\mathcal{H}$ (the upper half-plane) hence covered by a single coordinate patch, while G acts by global isometries which are “good” reparametrizations since—by definition—Eq. (2.4) is not spoiled by contact terms in the absence of 2d gravitational anomalies.¹⁵

We know from the discussion of the c -ghost zero-modes in Chap. 1 that *global* holomorphic vector fields on compact surfaces Σ generate global conformal motions. Then on a *compact* Σ there are only finitely many holomorphic vector fields, i.e.

$$\dim_{\mathbb{C}} H^0(\Sigma, T\Sigma) < \infty, \quad (2.62)$$

in fact (for Σ connected)

$$\dim_{\mathbb{C}} H^0(\Sigma, T\Sigma) \leq 3 \quad (2.63)$$

with equality iff $\Sigma \cong S^2 \equiv \mathbb{P}^1$; cf. discussion around Eq. (2.27).

Lie Algebras of Meromorphic Vector Fields

In 2d CFT we are also interested in holomorphic vector fields which do not generate global conformal automorphisms. We consider “punctured” surfaces of the form

$$\Sigma = \bar{\Sigma} \setminus D \quad (2.64)$$

where $\bar{\Sigma}$ is a compact Riemann surface and D a finite collection of points $\{p_i\} \subset \bar{\Sigma}$ or, in technically more precise terms, a simple effective divisor¹⁶

$$D = \sum_i p_i, \quad \text{supp } D = \{p_i\}. \quad (2.65)$$

We allow the vector fields $v(z)$ to have poles (only) at the “punctures” in $\text{supp } D$ ¹⁷ and consider the Lie algebra $\mathfrak{m}(\Sigma)$ of all such *meromorphic* vector fields endowed with their natural Lie bracket

¹⁵ Indeed, the isometry groups of the three simply connected Riemann surfaces S^2 , $\mathbb{C}$, and $\mathcal{H}$ endowed with a metric of constant curvature are groups of Möbius transformations (see Eq. (2.118)), so they satisfy the differential constraint which defines “good” maps $z \mapsto z'$.

¹⁶ For the rudiments of *divisors* on Riemannian surfaces ($\equiv$ complex curves), see BOX 2.2.

¹⁷ Technically, we identify the vector fields $v(z)$ with the elements of $H^0(\Sigma, T\Sigma \otimes \mathcal{O}(\sum_i n_i p_i))$ for $n_i \in \mathbb{Z}$; cf. BOX 2.2.

$$[v(z), w(z)] \equiv v(z) \partial_z w(z) - w(z) \partial_z v(z). \quad (2.66)$$

A vector field $v(z) \in \mathfrak{m}(\Sigma)$ is called *global* if it extends holomorphically to the compactification $\overline{\Sigma}$, i.e. if it generates a global conformal motion of $\overline{\Sigma}$. A global vector field $v(z)$ is automatically complete by **Theorem 2.1**. By the Riemann–Roch theorem (cf. Sect. 1.6.1), the dimension of the space of meromorphic vector fields $v(z)$ with poles of order at most $n_i \geq 0$ at $p_i \in \text{supp } D$ is

$$\dim H^0\left(\overline{\Sigma}, K^{-1} \otimes \mathcal{O}\left(\sum_i n_i p_i\right)\right) \geq 3 - 3g + \sum_i n_i, \quad (2.67)$$

with equality if $2 - 2g + \sum_i n_i > 0$ (see **BOX 2.2**). Therefore, *if we have at least one puncture on Σ , i.e. $\text{supp } D \neq \emptyset$, the Lie algebra $\mathfrak{m}(\Sigma)$ of meromorphic vector fields with poles in $\text{supp } D$ is infinite-dimensional.*

Meromorphic vector fields $v(z) \in \mathfrak{m}(\Sigma)$, which are not global (i.e. with $v(z) \notin \text{aut}(\overline{\Sigma})$), do *not* generate symmetries of Σ . However the corresponding chiral currents

$$T^v(z) \equiv T(z) v(z) \quad (2.68)$$

are conserved, $\bar{\partial} T^v(z) = 0$, and hence yield valid Ward identities that the quantum amplitudes on the punctured surface Σ should obey. For $D \neq 0$ we have infinitely many such meromorphic fields, spanning the infinite-dimensional Lie algebra $\mathfrak{m}(\Sigma)$, and an infinite set of Ward identities for the correlations which (in principle) allow us to compute all n -point functions.

It may look counterintuitive that we have conserved currents and Ward identities, but not associated symmetries. It should be in this way, given that the conserved currents generate an infinite-dimensional Lie algebra $\mathfrak{m}(\Sigma)$, while the Lie group of symmetries is always finite-dimensional. The point is that the functor

$$\left[\begin{array}{c} \text{finite-dimensional} \\ \text{Lie algebras} \end{array} \right] \xrightarrow{\exp} \left[\begin{array}{c} \text{simply connected, finite} \\ \text{dimensional Lie groups} \end{array} \right], \quad (2.69)$$

given by the exponential map, does not extend to infinite-dimensional Lie algebras.

Example: polynomial vector fields

We take $\Sigma = \mathbb{C} \equiv \mathbb{P}^1 \setminus \{\infty\}$ and consider the polynomial vector fields $z^{\ell+1} \partial_z$ with $\ell \geq 1$. Formally the corresponding would-be finite transformation of parameter t is the “map” $f_\ell(\cdot, t): \mathbb{C} \rightarrow \mathbb{C}$

$$z \mapsto f_\ell(z, t) \equiv \exp\left(t z^{\ell+1} \partial_z\right) z, \quad (2.70)$$

satisfying the ODE and initial condition

$$\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial(\frac{1}{\ell^2 t})} \right) f_\ell(z, t) = 0, \quad f_\ell(z, 0) = z, \quad (2.71)$$

whose solution is

$$f_\ell(z, t) = \frac{z}{(1 - \ell t z^\ell)^{1/\ell}}. \quad (2.72)$$

For $\ell > 1$ this function has branch cuts, and is not a univalued holomorphic map $\mathbb{C} \rightarrow \mathbb{C}$. However its branch cuts start at the ℓ th roots of $1/(\ell t)$, so the map is *locally* well-defined (in say) the domain

$$D_\ell(t) = \{z \in \mathbb{C} : |z| < |\ell t|^{-1/\ell}\} \subset \mathbb{C}. \quad (2.73)$$

As $t \rightarrow 0$ the domain $D_\ell(t)$ covers the full plane. Thus, while *finite* transformations do not make sense for $\ell \geq 2$, infinitesimal ones are well-defined, and they suffice to produce conserved currents.

Radial Quantization

We specialize the radial quantization of Sect. 2.2.2 to 2d. Free propagation of a closed string corresponds to a cylindrical world-sheet

$$\text{Cy} = S^1 \times \mathbb{R} \quad (2.74)$$

where $\mathbb{R}$ is thought of as the Euclidean time. Let $w = \tau + i\sigma$ be the coordinate on the cylinder, with σ a periodic real coordinate of period 2π along S^1 and τ the coordinate for the factor $\mathbb{R}$. The cylinder Cy is conformally equivalent to the punctured plane $\mathbb{C}^\times$ via the holomorphic map

$$\text{Cy} \rightarrow \mathbb{C}^\times \equiv \mathbb{P}^1 \setminus \{0, \infty\}, \quad w \mapsto z \equiv \exp(w). \quad (2.75)$$

The infinite past $\tau \rightarrow -\infty$ corresponds to the origin $z = 0$, while the infinite future $\tau \rightarrow +\infty$ to $z = \infty$ on the Riemann sphere $\mathbb{P}^1$. Equal-time surfaces $\tau = \text{const.}$ are mapped into circles centered at the origin, $|z| = e^\tau$, and the Hamiltonian producing translations on the cylinder becomes the scaling operator.¹⁸ As discussed in Sect. 2.2.2 radial quantization implies the state-operator correspondence ι .

A basis of meromorphic vector fields with poles in $\{0, \infty\}$ is

$$\{z^{n+1} \partial_z\}_{n \in \mathbb{Z}}. \quad (2.76)$$

Very naively they generate holomorphic “reparametrizations” of the punctured plane

$$z \mapsto z' \equiv z + \epsilon z^{n+1} + O(\epsilon^2) \quad (2.77)$$

but, as illustrated in the **example** above, for all but finitely many $n \in \mathbb{Z}$ these are merely *local* reparametrizations defined only in suitable sub-domains of $\mathbb{C}^\times$. As shown in **BOX 1.11**, the global holomorphic vector fields on $\mathbb{P}^1$ are

$$\{\partial_z, z \partial_z, z^2 \partial_z\} \quad (2.78)$$

which generate the three-dimensional Lie algebra $\mathfrak{sl}(2, \mathbb{C})$.

¹⁸ More precisely, the identification holds for the adjoint action of the two Hamiltonians; there is a c -number shift between the two Hamiltonians; cf. Eq. (2.123).

BOX 2.2 - Holomorphic line bundles versus divisors

Divisors yield an alternative language for holomorphic line bundles over a complex manifold. We limit ourselves to manifolds of complex dimension 1, i.e. to Riemann surfaces.

A (Weil) *divisor* D on Σ is an element of the free Abelian group over the points of Σ , i.e. a finite formal sum $\sum_i n_i p_i$ where $p_i \in \Sigma$ and $n_i \in \mathbb{Z}$. The divisor is called *effective* if $n_i \geq 0$ for all i . The set of points $\{p_i : n_i \neq 0\} \equiv \text{supp } D$ is the *support* of D . For Σ smooth we have a different presentation of D . Let $\cup_i U_i = \Sigma$ be an open cover fine enough so that each U_i contains just one point $p_i \in \text{supp } D$ while $p_j \notin U_i$ for $i \neq j$. Let z_i be a local coordinate in U_i centered at p_i . We identify the divisor $D = \sum_i n_i p_i$ with the set of local meromorphic functions $\{\psi_i \equiv z_i^{n_i} : U_i \rightarrow \mathbb{C}\}$. By construction, if $U_i \cap U_j \neq \emptyset$ ($i \neq j$), the function $\psi_i|_{U_i \cap U_j} : U_i \cap U_j \rightarrow \mathbb{C}$ is holomorphic without zeros nor poles. Conversely, given an open cover $\cup_i U_i = \Sigma$ and *non-zero* local meromorphic functions $\psi_i : U_i \rightarrow \mathbb{C}$ such that the restrictions $\psi_i|_{U_i \cap U_j}$ have no zeros nor poles, we have a divisor

$$D = \sum_i \left(\sum_{p \in U_i} \text{ord}_p(\psi_i) p \right).$$

(U_i, ψ_i) and (V_α, ϕ_α) define the same divisor if, after restricting the local meromorphic functions to the open sets of a common refinement $\{W_\alpha\}$ of the two covers $\{U_i\}$ and $\{V_\alpha\}$, $\psi_\alpha/\phi_\alpha = h_\alpha$ is a nowhere vanishing holomorphic function in W_α . For readers familiar with sheaf theory, a divisor is a global section of the quotient sheaf $\mathcal{M}^*/\mathcal{O}^*$, where $\mathcal{M}^*$ is the sheaf of germs of invertible meromorphic functions and $\mathcal{O}^*$ the sheaf of germs of nowhere vanishing holomorphic functions. Since $H^1(\Sigma, \mathcal{M}^*) = 0$, the exact sequence

$$1 \rightarrow \mathcal{O}^* \rightarrow \mathcal{M}^* \rightarrow \mathcal{M}^*/\mathcal{O}^* \rightarrow 1$$

yields

$$\text{Pic}(\Sigma) \equiv H^1(\Sigma, \mathcal{O}^*) = H^0(\Sigma, \mathcal{M}^*/\mathcal{O}^*)/[H^0(\Sigma, \mathcal{M}^*)/\mathbb{C}^\times], \quad \spadesuit$$

i.e. the group of isoclasses of holomorphic line bundles on Σ , $\text{Pic}(\Sigma)$, is isomorphic to the quotient group of the divisors, modulo the divisors of *global* meromorphic functions.

Let us rephrase this statement in more elementary terms. Given a divisor $D \equiv \{U_i, \psi_i\}$ on the non-empty intersections $U_i \cap U_j$, we have the nowhere vanishing holomorphic functions $\lambda_{ij} = \psi_i/\psi_j$ taking values in $GL(1, \mathbb{C})$; the collection $\{\lambda_{ij}\}$ satisfies the cocycle condition (see BOX 1.6), so it defines a line bundle which we write $\mathcal{O}(D)$. The collection of local holomorphic functions $\{\psi_i\}$ is then a section of the line bundle $\mathcal{O}(D)$, since $\psi_i = \lambda_{ij}\psi_j$ on $U_i \cap U_j$. This leads to a correspondence from divisors to line bundles; two sets of local functions, $\{\psi_i\}$ and $\{\phi_i\}$ which define the same divisor D clearly, are the same section of $\mathcal{O}(D)$ in different local trivializations. It remains to see when two divisors D, D' define isomorphic line bundles $\mathcal{O}(D) \simeq \mathcal{O}(D')$. A divisor D is called *principal* if there is a (non-zero) global meromorphic function f on Σ such that

$$D = \sum_{p \in \Sigma} \text{ord}_p(f) p.$$

We claim that $\mathcal{O}(D)$ is the trivial line bundle if and only if D is principal. Indeed, a line bundle is trivial iff in some holomorphic trivialization a section is a global holomorphic function. We say that two divisors D, D' are *linearly equivalent* $D \sim D'$ if $D - D'$ is principal. Then we have (cf. Eq. (♠) which is the same statement in sheaf language).

Theorem *The (multiplicative) Picard group $\text{Pic}(\Sigma)$ is isomorphic to the (additive) group of divisors modulo linear equivalence through the map $\mathcal{O}(D) \mapsto D$.*

We define the degree of a divisor to be the degree of the corresponding line bundle

$$\deg D = \int_{\Sigma} c_1(\mathcal{O}(D)) \in \mathbb{Z}.$$

Lemma One has $\deg(\sum_i n_i p_i) = \sum_i n_i \in \mathbb{Z}$. Principal divisors have degree zero.

From the Riemann–Roch theorem, and using the standard notation $\mathcal{L}(D) \equiv \mathcal{L} \otimes \mathcal{O}(D)$ for all line bundles $\mathcal{L}$ and divisors D , we have (for Σ compact)

$$\dim H^0(\Sigma, \mathcal{O}(D)) - \dim H^0(\Sigma, K(-D)) = 1 - g + \deg D.$$

$H^0(\Sigma, \mathcal{O}(D)) \neq 0$ iff D is linearly equivalent to an effective divisor (the divisor of a holomorphic section being effective); thus $H^0(\Sigma, \mathcal{O}(D)) = 0$ if the degree of D is negative.

Corollary The dimension of the space of holomorphic sections of a line bundle $\mathcal{L}$ with orders at least $\{n_i\}$ at given points $\{p_i\}$ is $\dim H^0(\Sigma, \mathcal{L}(-D))$ where $D = \sum_i n_i p_i$.

Pf Let $D = (U_i, \psi_i)$ and s be a holomorphic section of $\mathcal{L}$ with $\text{ord}_{p_i} s \geq n_i$. The set of local functions $\{s|_{U_i}/\psi_i\}$ is a holomorphic section of $\mathcal{L}(-D)$ and all such sections are of this form.

2.3.1 Primary Fields

The basic objects of 2d CFT are the conformal local operators, also called *primary fields* $\phi(z, \bar{z})$. They are defined by the property that under a *local* conformal transformation $z \mapsto z' \equiv z'(z)$ the primary fields transform as tensors

$$\phi(z, \bar{z}) \rightarrow \phi'(z', \bar{z}') \equiv \left(\frac{\partial z'}{\partial z} \right)^{-h} \left(\frac{\partial \bar{z}'}{\partial \bar{z}} \right)^{-\tilde{h}} \phi(z, \bar{z}), \quad (2.79)$$

that is, as $(h, \tilde{h})$ -differentials

$$\phi'(z', \bar{z}')(dz')^h (d\bar{z}')^{\tilde{h}} = \phi(z, \bar{z})(dz)^h (d\bar{z})^{\tilde{h}}, \quad (2.80)$$

where the real numbers $h, \tilde{h}$ are called *conformal weights* (or Virasoro weights). On a compact Riemann surface Σ , the primary field ϕ is a section of the smooth line bundle¹⁹

$$K^h \otimes \bar{K}^{\tilde{h}}. \quad (2.81)$$

Working locally in a coordinate patch, we may formally take $h, \tilde{h}$ to be real numbers (subject to the unitary bounds to be discussed momentarily). However the global existence of the line bundle (2.81) on an arbitrary surface Σ gives restrictions. Since $\bar{K} \simeq K^{-1}$ (in the smooth sense²⁰), we have

¹⁹ As always K stands for the *canonical line bundle*; see BOX 1.6.

²⁰ The C^∞ -isomorphism $\bar{K} \simeq K^{-1}$ is given by any Kähler metric: $v_{\bar{z}} = g_{z\bar{z}} v^z$.

$$K^h \otimes \bar{K}^{\tilde{h}} \simeq K^{h-\tilde{h}} \quad (\text{as } C^\infty \text{ bundles !}), \quad (2.82)$$

so the requirement for ϕ to be globally well-defined on Σ is²¹

$$\text{conformal spin of } \phi \stackrel{\text{def}}{=} h - \tilde{h} \in \tfrac{1}{2}\mathbb{Z}. \quad (2.83)$$

As we shall see in a moment, (2.83) is the condition for the field ϕ to be local *with respect to itself* which is the minimal requirement for a genuine local operator. $h - \tilde{h}$ is then either integral or half-integral. To define the operator ϕ in the second case, we need to choose a *spin-structure* on Σ as described in Chap. 1; cf. **BOX 1.10**.

Purely holomorphic (resp. anti-holomorphic) primary fields are called *chiral*: they have $\tilde{h} = 0$ (resp. $h = 0$). The combination

$$h + \tilde{h} \equiv \Delta \quad (2.84)$$

is the *scaling dimension* of the primary field ϕ , i.e. the eigenvalue of the dilation operator (the Hamiltonian of radial quantization) on the state $|\phi\rangle$ which corresponds to ϕ under the state-operator map ι . The other combination, $h - \tilde{h}$, the *conformal spin*, is the eigenvalue of the rotation operator around the origin on the state $|\phi\rangle$. A state $|\phi\rangle$ is called *primary* if it corresponds via ι to a primary operator.

Local fields which are not primary are called *descendant*. They are not independent operators of the CFT but rather derivatives of the primary ones, as we shall see later. The CFT can be fully reconstructed from the correlators of primary operators.

Convention/Warning To write more compact formulae, most of the time we shall write only the left-moving (holomorphic) part of each expression; we leave implicit the factor which refers to the right-moving (anti-holomorphic) side which looks exactly the same up to putting *tildes* and *bars* over the relevant symbols. For example, Eq. (2.80) will be written simply as

$$\phi'(z')(dz')^h = \phi(z)(dz)^h \quad (2.85)$$

leaving implicit the anti-holomorphic side of the expressions. When there is danger of confusion, we shall restore the right-movers and write the formula in full.

The Cylinder versus the Plane

Free propagation of a closed oriented string is described by the 2d CFT on a world-sheet Σ which is an infinite cylinder $S^1 \times \mathbb{R}$. The cylinder is conformally equivalent to the punctured plane $\mathbb{C}^\times$ via the biholomorphic map (2.75) which acts on a conformal primary $\phi(z, \bar{z})$ as (cf. Eq. (2.80))

$$\phi(z, \bar{z})_{\text{plane}} = z^{-h} \bar{z}^{-\tilde{h}} \phi(w, \bar{w})_{\text{cyl}}. \quad (2.86)$$

²¹ Recall (**BOX 1.6**) that the Chern class ($\equiv$ degree) of the bundle $K^{h-\tilde{h}}$ on a surface of genus g is $2(h - \tilde{h})(g - 1)$ which is integral for all g iff (2.83) holds.

Note that a primary operator which is univalued on the cylinder is also univalued on the plane if and only if

$$h - \tilde{h} \in \mathbb{Z}. \quad (2.87)$$

On the contrary, when $h - \tilde{h} \in \mathbb{Z} + \frac{1}{2}$, an operator univalued in one of the two spaces is double-valued in the other one: its two values differ by a sign. To define a field ϕ with *half-integral* $h - \tilde{h}$, we need to specify a spin-structure on Σ (cf. **BOX 1.10**). The cylinder (and hence $\mathbb{C}^\times$) has *two* distinct spin-structures: the operator $\phi(w)$ may be either *periodic* or *anti-periodic* under the shift $w \rightarrow w + 2\pi i$. In the first case $\phi(z)$ will be doubly valued in the punctured plane (it changes sign under a 2π rotation around the origin), while in the second case it would be univalued in $\mathbb{C}^\times$.

Mode Expansions

(Warning: In this paragraph we only write the left-moving part of the expressions).

Let $\phi(z)$ be a complex primary operator of weight h . We assume that $\phi(z)$ obeys a generalized periodic boundary condition on the cylinder $S^1 \times \mathbb{R}$ of the form

$$\phi(w + 2\pi i) = e^{2\pi i x} \phi(w), \quad 0 \leq x < 1. \quad (2.88)$$

The mode expansion on the cylinder (Fourier series) then takes the form

$$\phi(w)_{\text{cyl}} = \sum_{n \in \mathbb{Z} - x} \phi_n e^{-nw}, \quad w \sim w + 2\pi i, \quad (2.89)$$

for some quantum operators ϕ_n . Using Eq. (2.86) we transform this formula into the mode expansion on the plane (Laurent series)

$$\phi(z)_{\text{plane}} = \sum_{n \in \mathbb{Z} - x} \frac{\phi_n}{z^{n+h}}. \quad (2.90)$$

For $\phi(z)$ to be *single-valued* on the complex plane, we need $n + h \in \mathbb{Z}$, i.e.

$$h - x \in \mathbb{Z}. \quad (2.91)$$

In this case the mode operators ϕ_n are given by Cauchy's formula

$$\phi_n = \oint_{C_0} \frac{dz}{2\pi i z} z^{n+h} \phi(z) \quad (2.92)$$

where the integral is along a contour C_0 encircling the origin.

2.3.2 The Virasoro Algebra

The conserved charge associated with the current $T(z) v(z)$ and closed contour C is

$$Q^v(C) = \oint_C \frac{dz}{2\pi i} v(z) T(z). \quad (2.93)$$

In radial quantization we usually take C to be an equal-time curve, i.e. a circle centered in the origin. If C and C' are two such circles of radii $r < r'$, and there is no operator insertion in between them, we have $Q^v(C) = Q^v(C')$ and we write the corresponding conserved charge simply Q^v . Suppose a local operator $O(w)$ is inserted at a point w in the annulus $r < |z| < r'$ bounded by the contours C and C' . In radial quantization the quantum operators are time-ordered in radial time; hence

$$\begin{aligned} [Q^v, O(w)] &= Q^v O(w) - O(w) Q^v = \\ &= \oint_{C'} \frac{dz}{2\pi i} v(z) T(z) O(w) - \oint_C \frac{dz}{2\pi i} v(z) T(z) O(w) = \\ &= \oint_{C_w} \frac{dz}{2\pi i} v(z) T(z) O(w) \stackrel{\text{def}}{=} \delta_v O(w) \end{aligned} \quad (2.94)$$

where C_w is a small loop enclosing the point w and we used that the cycle $C' - C$ is homologous to C_w . From this contour manipulation, it is clear that the commutators of local operators are determined by the *singular part* of their OPEs. The last equality in Eq. (2.94) is the *definition* of the infinitesimal variation $\delta_v O(w)$ of the local operator $O(w)$ along the holomorphic vector field $v(z)$. The charge Q^v generates the infinitesimal conformal transformation

$$z \rightarrow z' = z + \epsilon v(z) \quad (2.95)$$

which is well-defined as discussed around Eq. (2.68). Expanding (2.85), we get the variation of a primary field $\phi(z)$ under the infinitesimal conformal motion (2.95)

$$\begin{aligned} \epsilon \delta_v \phi(z) + O(\epsilon^2) &= \phi(z') (dz'/dz)^h - \phi(z) = \\ &= \phi(z + \epsilon v)(1 + \epsilon \partial v)^h - \phi(z) = \epsilon (v(z) \partial + h \partial v) \phi(z) + O(\epsilon^2). \end{aligned} \quad (2.96)$$

Then for a primary field $\phi(w)$, we have

$$\oint_{C_w} \frac{dz}{2\pi i} v(z) T(z) \phi(w) = [Q^v, \phi(w)] \equiv \delta_v \phi(w) = (h \partial v(w) + v(w) \partial) \phi(w). \quad (2.97)$$

By the Cauchy residue theorem,

$$\oint_{C_w} \frac{dz}{2\pi i} \frac{f(z)}{(z-w)^n} = \frac{1}{(n-1)!} f^{(n-1)}(w) \quad (f(z) \text{ holomorphic}), \quad (2.98)$$

Eq. (2.97) is equivalent to the following OPE between the energy–momentum-tensor $T(z)$ and the conformal primary $\phi(w)$

$$T(z) \phi(w) = \frac{h \phi(w)}{(z-w)^2} + \frac{\partial \phi(w)}{(z-w)} + \text{non singular as } z \rightarrow w. \quad (2.99)$$

The OPE (2.99) is equivalent to Eq. (2.85) and may be taken as *the definition* of primary operator. A local operator $\phi(w)$ of definite spin/dimension is primary iff its OPE with the energy–momentum tensor $T(z)$ has at worst poles of order 2 when $z \rightarrow w$. The weight h of $\phi(w)$ is the coefficient of the order-2 pole in its OPE with $T(z)$.

Acting on local fields, the Lie algebra of conformal motions takes the form

$$[\delta_{v_1}, \delta_{v_2}] \phi(z) = \delta_{[v_1, v_2]} \phi(z), \quad (2.100)$$

where $[v_1, v_2]$ is the Lie bracket of vector fields. Explicitly this equation reads²²

$$\begin{aligned} \oint_{C'_z} \frac{dw}{2\pi i} v_1(w) T(w) \oint_{C_z} \frac{dy}{2\pi i} v_2(y) T(y) \phi(z) - \oint_{C'_z} \frac{dw}{2\pi i} v_2(w) T(w) \oint_{C_z} \frac{dy}{2\pi i} v_1(y) T(y) \phi(z) = \\ = \oint_{C_z} \frac{dw}{2\pi i} [v_1(w), v_2(w)] T(w) \phi(z), \end{aligned} \quad (2.101)$$

which, in view of (2.98), implies the OPE

$$T(z) T(w) = \frac{c}{2(z-w)^4} + \frac{2 T(w)}{(z-w)^2} + \frac{\partial T(w)}{(z-w)} + \text{regular} \quad (2.102)$$

where c is a real parameter (called the *central charge*) which is not fixed by Eq. (2.101): indeed Eq. (2.101) fixes all singular terms in the OPE except the one proportional to the identity operator 1 which commutes with all local operators $\phi(z)$. The functional dependence on z and w of the term proportional to the identity is fixed by translation and scale symmetries up to the overall normalization given by the constant c . The OPE of the energy–momentum tensor with itself, Eq. (2.102), is the most important equation in this chapter.

Classically $c = 0$, but quantum mechanically $c \neq 0$. In **BOX 2.3**, we show that c is a precise measure of the Weyl anomaly of our CFT when defined on a two-dimensional spacetime Σ with a background metric g_{ab} of scalar curvature R :

$$T^a_a = -\frac{c}{12} R. \quad (2.103)$$

The OPE (2.102) says that $T(z)$ has scaling dimension 2 (as expected) but it is *not* a primary operator. In fact $T(z)$ is a descendant of the identity operator. The infinitesimal conformal transformation of $T(z)$ then has an additional term

²² The closed contour C'_z contains the closed contour C_z which in turn encircles the point z .

$$\delta_v T(z) = \oint \frac{dw}{2\pi i} v(w) T(w) T(z) = 2 \partial v(z) T(z) + v(z) \partial T(z) + \frac{c}{12} \partial^3 v(z). \quad (2.104)$$

Mode Expansions of $T(z)$

The mode expansion of the energy–momentum tensor has the form

$$T(z) = \sum_{n \in \mathbb{Z}} \frac{L_n}{z^{n+2}}, \quad L_n = \oint \frac{dz}{2\pi i} z^{n+1} T(z) \quad (2.105)$$

$$\tilde{T}(\bar{z}) = \sum_{n \in \mathbb{Z}} \frac{\tilde{L}_n}{\bar{z}^{n+2}}, \quad \tilde{L}_n = \oint \frac{d\bar{z}}{-2\pi i} \bar{z}^{n+1} \tilde{T}(\bar{z}) \quad (2.106)$$

with the Hermiticity conditions

$$(L_n)^\dagger = L_{-n}, \quad (\tilde{L}_n)^\dagger = \tilde{L}_{-n}, \quad (2.107)$$

which follow from the reality of the energy–momentum tensor after analytic continuation to Minkowski signature.

Let us compute the commutator $[L_m, L_n]$ using Cauchy residues

$$\begin{aligned} [L_m, L_n] &= \oint_{C_0} \frac{dw}{2\pi i} \oint_{C_w} \frac{dz}{2\pi i} z^{m+1} w^{n+1} \left[\frac{c}{2(z-w)^4} + \frac{2T(w)}{(z-w)^2} + \frac{\partial T(w)}{(z-w)} \right] \\ &= \frac{c}{12} m(m^2 - 1) \delta_{m+n,0} + (m-n) L_{m+n}. \end{aligned} \quad (2.108)$$

The Lie algebra of the modes of $T(z)$

$$[L_m, L_n] = \frac{c}{12} m(m^2 - 1) \delta_{m+n,0} + (m-n) L_{m+n}, \quad (m, n \in \mathbb{Z}) \quad (2.109)$$

is called the *Virasoro algebra*. It is a central extension of the classical *loop algebra*²³ (which is the case $c = 0$ of (2.109)). For this reason the constant c is called the *Virasoro central charge*. Equation (2.109) has the same physical content as the OPE between two energy–momentum tensors, Eq. (2.102).

In a CFT there is a second *right-moving* isomorphic copy of the Virasoro algebra with generators $\tilde{L}_n$; cf. Eq. (2.106). The two algebras commute

$$[L_m, \tilde{L}_n] = 0. \quad (2.110)$$

Claim 2.1 In a 2d CFT, the absence of 2d gravitational anomalies (i.e. $\mathbf{Diff}^+$ -invariance at the quantum level) requires the left and right Virasoro central charges to be equal

$$c = \tilde{c}. \quad (2.111)$$

²³ The loop algebra has a one-parameter family of non-trivial central extensions parametrized by c .

BOX 2.3 - Weyl anomaly versus Virasoro central charge c

Contact Terms In a 2d CFT we have $T_{\bar{z}z}(z) = 0$ as an *operator* equation. This implies that the two-point function of the trace part of the energy–momentum tensor vanishes

$$\langle T_{\bar{z}z}(z) T_{\bar{w}w}(w) \rangle_{S^2} = 0 \quad \star$$

when the points z, w are *distinct*; for $z = w$ we may have a *contact term*. In other words, the LHS of (★) is not zero but rather a distribution with support on $z = w = 0$, i.e. a finite-order derivative of the delta function $\delta^{(2)}(z - w)$. The derivative order must be 2 by scaling properties. We recall the Poincaré–Lelong formula [20]: in the distributional sense, in $\mathbb{C}$ we have the identity

$$\partial_{\bar{z}} \frac{1}{z^{n+1}} = (-1)^n \frac{2\pi}{n!} \partial_z^n \delta^{(2)}(z).$$

From the OPE (2.102) and $\langle T_{zz}(z) \rangle = 0$ (cf. Eq. (2.180)), we get

$$\langle T_{zz}(z) T_{ww}(w) \rangle_{S^2} = \frac{c}{2(z-w)^4}.$$

Since (by assumption) we are quantizing the theory while preserving the local **Diff** invariance, there is no contact term spoiling the Ward identity of the energy–momentum tensor, that is,

$$\partial_{\bar{z}} \langle T_{zz}(z) O(w) \rangle_{S^2} + \partial_z \langle T_{\bar{z}\bar{z}}(z) O(w) \rangle_{S^2} = \langle \partial_w O(w) \rangle_{S^2} = 0$$

everywhere and for all $O(w)$. (The last equality is translational invariance). Therefore

$$\begin{aligned} \partial_z \partial_w \langle T_{\bar{z}\bar{z}}(z) T_{\bar{w}w}(w) \rangle_{S^2} &= \partial_z \partial_w \langle T_{zz}(z) T_{ww}(w) \rangle_{S^2} = \frac{c}{2} \partial_z \partial_w \frac{1}{(z-w)^4} = \\ &= -\frac{2\pi c}{2 \cdot 3!} \partial_z \partial_w^3 \delta^{(2)}(z-w) = -\frac{\pi c}{6} \partial_z \partial_w (\partial_z \partial_z \delta^{(2)}(z-w)). \end{aligned}$$

The LHS of (★) has support on $z = w$, and $\partial_z \partial_w$ is injective on such distributions; thus

$$\langle T_{\bar{z}\bar{z}}(z) T_{\bar{w}w}(w) \rangle_{S^2} = -\frac{\pi c}{6} \partial_z \partial_z \delta^{(2)}(z-w).$$

A non-zero contact term in the correlation functions of the trace $T_{\bar{z}z}(z)$ signals a *local* Weyl anomaly. We conclude that the Weyl anomaly is proportional to the Virasoro central charge c .

$\langle T_{\bar{z}\bar{z}} \rangle$ in a Metric Background Next we compute $\langle T_{\bar{z}\bar{z}}(z, \bar{z}) \rangle$ as a function of the metric background. We choose local coordinates so that $g_{z\bar{z}} = e^{2\phi}$. The Weyl variation of $\langle T_{\bar{z}\bar{z}}(z, \bar{z}) \rangle$ can be computed using Eq. (1.91)

$$\begin{aligned} \delta \langle T_{\bar{z}\bar{z}}(z, \bar{z}) \rangle &= -\frac{1}{2\pi} \int d^2w \delta\phi(\sigma) \langle T_{w\bar{w}}(w, \bar{w}) T_{\bar{z}\bar{z}}(z, \bar{z}) \rangle = \\ &= \frac{c}{12} \partial_z \partial_{\bar{z}} \delta\phi(z, \bar{z}) = -\frac{c}{12} g_{z\bar{z}} \delta R \end{aligned}$$

whose integral is the equation we used in Chap. 1 to determine the critical dimension

$$\langle T^a{}_a(\sigma) \rangle = -\frac{c}{12} R + \text{const.}$$

Proof In **BOX 2.3**, it is shown that if there is a quantization-preserving $\mathbf{Diff}^+$ -invariance one has

$$\langle T_{\bar{z}z}(z) T_{\bar{w}w}(w) \rangle_{S^2} = -\frac{\pi c}{6} \partial_{\bar{z}} \partial_z \delta^{(2)}(z - w). \quad (2.112)$$

The LHS is invariant under $\text{left} \leftrightarrow \text{right}$, and so must be the RHS. This requires $c = \tilde{c}$. $\square$

Note 2.2 The condition $c = \tilde{c}$ guarantees the absence of *local* $\mathbf{Diff}^+$ -anomalies, i.e. invariance under “small” diffeomorphisms homotopic to the identity. A model with $c = \tilde{c}$ may still suffer *global* $\mathbf{Diff}^+$ -anomalies, i.e. not be invariant under “large” diffeomorphisms. We discuss global $\mathbf{Diff}^+$ -anomalies in Sect. 2.3.7 and in Chap. 5.

Action of L_n on Primary Fields

The L_n ’s generate infinitesimal conformal motions: indeed, comparing Eq. (2.93) with the second (2.105) we see that the n th mode L_n is just the charge $Q^{z^{n+1}}$ generating the infinitesimal conformal motion $\delta z = z^{n+1}$. Specializing Eq. (2.97) to $v(z) = z^{n+1}$, for $\phi(z)$ primary we get

$$[L_n, \phi(z)] \equiv [Q^{z^{n+1}}, \phi(z)] = \left(z^{n+1} \partial + (n+1)h z^n \right) \phi(z), \quad (2.113)$$

or, in terms of modes of the primary $\phi(z)$ (cf. Eq. (2.90)),

$$[L_n, \phi_m] = \left(n(h-1) - m \right) \phi_{m+n}. \quad (2.114)$$

We conclude that the Hilbert space $\mathbf{H}_{S^1}$ of a 2d CFT carries *two* commuting representations of the Virasoro algebra (2.109): the *left*- and *right-moving* ones. Besides the adjoint representation (2.113), the algebra $\mathfrak{A}$ of local operators carries a second Virasoro representation, written as $L_n \cdot \phi(z)$, via the rule $|L_n \cdot \phi\rangle = L_n |\phi\rangle$.

The $\mathfrak{sl}(2, \mathbb{C})$ Subalgebra

The Virasoro generators L_{-1}, L_0, L_{+1} generate a finite-dimensional Lie algebra isomorphic to $\mathfrak{sl}(2, \mathbb{C}) \simeq \mathfrak{so}(3, \mathbb{C})$,

$$[L_0, L_{\pm 1}] = \mp L_{\mp 1}, \quad [L_1, L_{-1}] = 2L_0, \quad (2.115)$$

i.e. to the Lie algebra of the global conformal automorphism group²⁴ $PSL(2, \mathbb{C})$ of the sphere $\mathbb{P}^1$. The corresponding infinitesimal motions have the form $z + \delta z$ with

$$\delta z \equiv v(z) = a + bz + cz^2; \quad (2.116)$$

cf. **BOX 1.11**. The *finite* $SL(2, \mathbb{C})$ transformations are the projective automorphisms of $\mathbb{P}^1 \simeq S^2$. The point $(w_1 : w_2) \in \mathbb{P}^1$ corresponds to the point $z \equiv w_1/w_2$ in the Riemann sphere $\mathbb{C} \cup \{\infty\} \simeq \mathbb{P}^1$, and a $PSL(2, \mathbb{C})$ projective transformation acts as

$$\begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \rightarrow \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \quad \text{with} \quad ad - bc = 1, \quad (2.117)$$

²⁴ The projective group $PSL(2, \mathbb{C})$ is the quotient of $SL(2, \mathbb{C})$ by its center $\mathbb{Z}_2$.

that is, as the fractional linear map

$$z \rightarrow z' \equiv \frac{w'_1}{w'_2} = \frac{aw_1 + bw_2}{cw_1 + dw_2} \equiv \frac{az + b}{cz + d} \quad (2.118)$$

a.k.a. a *Möbius transformation*. By general theory, the automorphisms of the Riemann sphere $\mathbb{P}^1 \rightarrow \mathbb{P}^1$ are the meromorphic functions of degree 1, i.e. with one zero and one pole. It is easy to check that they are precisely the Möbius transformations (2.118), i.e. the ratio of two polynomials of degree at most 1.

2.3.3 Finite Conformal Transformation of $T(z)$

From Eq. (2.104) one sees that the transformation of the energy–momentum tensor under a finite conformal transformation $z \mapsto w(z)$ is

$$T(w)(dw)^2 = \left(T(z) - \frac{c}{12} \{w, z\} \right) (dz)^2, \quad (2.119)$$

where $\{w, z\}$ is the *Schwarzian derivative*²⁵

$$\{w, z\} \stackrel{\text{def}}{=} \frac{w'''}{w'} - \frac{3}{2} \left(\frac{w''}{w'} \right)^2. \quad (2.120)$$

Exercise 2.4 Prove Eqs. (2.119), (2.120) and show that the Schwarzian derivative vanishes if and only if $z \mapsto w$ is a Möbius map.

We apply these formulae to the conformal map from the cylinder to the punctured plane $w \mapsto z = e^w$:

$$\{z, w\} = \frac{e^w}{e^w} - \frac{3}{2} \frac{e^{2w}}{e^{2w}} = -\frac{1}{2} \Rightarrow T(z)_{\text{plane}} = \frac{1}{z^2} \left(T(w)_{\text{cyl}} + \frac{c}{24} \right) \quad (2.121)$$

so, passing to modes,

$$(L_m)_{\text{cyl}} \equiv \int_0^{2\pi i} \frac{dw}{2\pi i} T(w)_{\text{cyl}} e^{mw} = L_m - \frac{c}{24} \delta_{m,0}, \quad (2.122)$$

where L_m are the modes in the plane; cf. (2.105). This formula gives a precise relation between the scaling dimension $h + \tilde{h}$, i.e. the eigenvalue of the radial Hamiltonian $L_0 + \tilde{L}_0$, and the eigenvalue of the Hamiltonian in a cylinder of circumference 2π

$$H_{\text{cyl}} \equiv \int_0^{2\pi} \frac{d\sigma}{2\pi} T_{00}(\sigma, 0)_{\text{cyl}} = L_0 + \tilde{L}_0 - \frac{c + \tilde{c}}{24}. \quad (2.123)$$

²⁵ For the Schwarzian derivative and its relation to conformal transformations, see [26]. In Eq. (2.120) each prime stands for one derivation with respect to z .

The vacuum has $h = \tilde{h} = 0$, so the vacuum energy (a.k.a. Casimir energy) of our CFT quantized in a circle of length L is²⁶

$$E_{\text{Casimir}} = -\frac{\pi}{12L}(c + \tilde{c}); \quad (2.124)$$

cf. **BOX 1.2**. The global conformal motions of the torus,²⁷ seen as a finite cylinder of circumference 2π with identified boundary components, are then generated by

$$H_L \stackrel{\text{def}}{=} L_0 - c/24 \quad \text{and} \quad H_R \stackrel{\text{def}}{=} \tilde{L}_0 - \tilde{c}/24. \quad (2.125)$$

2.3.4 Representations on the Hilbert Space

As always, we write $|0\rangle$ for the state $\iota(1)$; cf. Eq. (2.44). The state $T(z)|0\rangle$ has no singularity as $z \rightarrow 0$ since in its path integral Schrödinger representation

$$\langle \Phi_0 | T(z) | 0 \rangle = \int_{\Phi|_{\partial B} = \Phi_0} [d\Phi] e^{-S[\Phi]} T(z) \quad (2.126)$$

there is no non-trivial operator inserted at the origin. Using Eq. (2.105)

$$\lim_{z \rightarrow 0} \sum_{n \in \mathbb{Z}} \frac{L_n}{z^{n+2}} |0\rangle = \text{regular} \quad \Rightarrow \quad L_n |0\rangle = 0 \text{ for } n \geq -1. \quad (2.127)$$

Indeed the L_n with $n \geq -1$ generate infinitesimal motions $z \rightarrow z + \epsilon z^{n+1}$ which are regular at the origin. The dual vacuum $\langle 0|$ is invariant under the conformal motions which are regular at the point $z = \infty$ on the Riemann sphere. We write $z = 1/w$ with w the good coordinate around the point ∞ . Taking into account the transformation of $T(z)$ under the Möbius transformation $z \rightarrow 1/z$, we get²⁸

$$\begin{aligned} \text{regular} &= \lim_{w \rightarrow 0} \langle 0 | T(w) = \lim_{w \rightarrow 0} \langle 0 | \sum_{n \in \mathbb{Z}} w^{n-2} L_n \Rightarrow \\ &\Rightarrow \langle 0 | L_n = 0 \text{ for } n < 2. \end{aligned} \quad (2.128)$$

Note 2.3 The generators of $SL(2, \mathbb{C})$, $L_{\pm 1}$, and L_0 leave invariant both vacua $|0\rangle$ and $\langle 0|$. This agrees with the geometry of global conformal motions of S^2 ; cf. **BOX 1.11**.

Claim 2.2 $|0\rangle$ and its Hermitian conjugate $\langle 0|$ are the *only* states invariant under $SL(2, \mathbb{C})$. The state $|0\rangle$ is called the $SL(2, \mathbb{C})$ -invariant state (or vacuum).

²⁶ Recall that, in order not to have 2d gravitational anomalies, we need $c = \tilde{c}$.

²⁷ Which are just translations $w \rightarrow w + a$ by the last clause in **Theorem 2.1**.

²⁸ Equations (2.127), (2.128) are Hermitian conjugate of each other.

Proof By the state-operator isomorphism, all states are of the form $\mathcal{O}(0)|0\rangle$ for some $\mathcal{O}(z)$. Then

$$0 = L_{-1}(\mathcal{O}(0)|0\rangle) = [L_{-1}, \mathcal{O}(0)]|0\rangle = (\partial\mathcal{O}(0))|0\rangle \quad (2.129)$$

and the identity 1 is the only local field whose derivative vanishes.²⁹ $\square$

More on Primary States

Let $\phi(z)$ be a primary operator of weight h . $\phi(z)|0\rangle$ is again regular as $z \rightarrow 0$; in view of Eq. (2.90) this yields

$$\phi_n|0\rangle = 0 \text{ for } n > -h \text{ and } \langle 0|\phi_n = 0 \text{ for } n < h. \quad (2.130)$$

We write the state-operator correspondence in terms of modes

$$|\phi\rangle = \lim_{z \rightarrow 0} \phi(z)|0\rangle = \phi_{-h}|0\rangle = \oint \frac{dz}{2\pi i z} \phi(z)|0\rangle, \quad (2.131)$$

and, more generally,

$$\phi(z)|0\rangle = e^{zL_{-1}}|\phi\rangle, \quad (2.132)$$

while

$$\langle \phi^\dagger | = \lim_{z \rightarrow \infty} z^{2h} \langle 0 | \phi(z)^\dagger = \langle 0 | (\phi^\dagger)_h. \quad (2.133)$$

We define the positive-definite Hermitian inner product³⁰

$$\langle \phi_i^\dagger | \phi_j \rangle \stackrel{\text{def}}{=} \lim_{\substack{z' \rightarrow 0 \\ z \rightarrow \infty}} z^{2h_i} \langle 0 | \phi_i^\dagger(z) \phi_j(z') | 0 \rangle. \quad (2.134)$$

Often we use a basis of primary operators which is orthonormal with respect to this Hermitian product. Let ϕ be a primary operator:

$$\begin{aligned} \text{for } n > 0 \quad L_n|\phi\rangle &= L_n\phi_{-h}|0\rangle = \\ &= [L_n, \phi_{-h}]|0\rangle = (n(h-1) + h)\phi_{n-h}|0\rangle = 0 \end{aligned} \quad (2.135)$$

$$\text{for } n = 0 \quad L_0|\phi\rangle = L_0\phi_{-h}|0\rangle = [L_0, \phi_{-h}]|0\rangle = h\phi_{-h}|0\rangle = h|\phi\rangle. \quad (2.136)$$

The two equations

$$L_n|\phi\rangle = 0 \text{ for } n > 0 \text{ and } L_0|\phi\rangle = h|\phi\rangle \quad (2.137)$$

²⁹ **Proof.** In QFT local fields commute at space-like separation. If $\mathcal{O}(z)$ is a local field with vanishing derivative and $\phi(w)$ any local field, we have $[\mathcal{O}(z), \phi(w)] = [\mathcal{O}(y), \phi(w)]$ where y is any arbitrarily chosen point. Choosing y to be space-like to w , we see that $\mathcal{O}(z)$ commutes with all local operators, so it is a central element of $\mathfrak{A}$. The axioms of QFT imply $\mathcal{Z}(\mathfrak{A}) = \mathbb{C} \cdot 1$; cf. **Theorem 4–5** of [27].

³⁰ **Caviat:** We have defined $\langle \phi | \equiv \langle 0 | \phi_h$ to be *linear* in the field ϕ not *anti-linear*. To get a Hermitian form we take by hand the Hermitian conjugate of the first argument. This is the opposite convention with respect to Quantum Mechanics, but it is the natural one in the present context.

may be taken as the definition of a *primary state* $|\phi\rangle$ of *weight* h . They are equivalent to saying that the OPE of $\phi(w)$ with $T(z)$ contains poles of order at most 2 while the coefficient of the double pole is h . Note that a state is primary iff

$$L_1|\phi\rangle = L_2|\phi\rangle = 0, \quad (2.138)$$

all other conditions (2.137) being implied by the Virasoro algebra (2.109). We have

$$L_0(L_{-n}|\phi\rangle) = (n+h)(L_{-n}|\phi\rangle) \quad n \geq 0, \quad (2.139)$$

i.e. the L_{-n} 's with $n > 0$ raise the eigenvalue of L_0 by n .

Quasi-Primary States

A state $|\psi\rangle$ which satisfies the first Eq. (2.138), $L_1|\psi\rangle = 0$, is called *quasi-primary*. The corresponding local operator $\psi(z)$ is also called quasi-primary. A quasi-primary state is primary iff, in addition, it satisfies the equation $L_2|\psi\rangle = 0$. Quasi-primary states are highest weight states for the finite-dimensional $\mathfrak{sl}(2, \mathbb{C})$ subalgebra of the Virasoro algebra. The state $|T\rangle$, hence the energy–momentum tensor $T(z)$, is quasi-primary

$$L_1|T\rangle = L_1L_{-2}|0\rangle = [L_1, L_{-2}]|0\rangle = 3L_{-1}|0\rangle = 0. \quad (2.140)$$

Verma Modules

L_0 spans the Cartan subalgebra of the Virasoro Lie algebra (2.109), while L_{-n} ($n > 0$) are raising operators. In the language of Lie representation theory [28], Eq. (2.137) just says that the primary states are *highest*³¹ weight vectors for the Virasoro algebra. The full Hilbert space is then obtained by acting with the operators L_{-n} ($n > 0$) on the highest weight states $\{|\phi_j\rangle\}$. By the Poincaré–Birkhoff–Witt theorem [29] the representation generated by acting with the L_n 's on $|\phi_j\rangle$ is spanned by the vectors

$$|\phi_j^{n_1, n_2, \dots, n_s}\rangle \stackrel{\text{def}}{=} L_{-n_1}L_{-n_2} \cdots L_{-n_s}|\phi_j\rangle, \quad n_1 \geq n_2 \geq \cdots \geq n_s > 0, \quad (2.141)$$

for all finite non-increasing sequences of positive integers n_i . One has

$$L_0|\phi_j^{n_1, n_2, \dots, n_s}\rangle = \left(h_j + \sum_k n_k\right)|\phi_j^{n_1, n_2, \dots, n_s}\rangle. \quad (2.142)$$

The representation generated by the states $|\phi_j^{n_1, n_2, \dots, n_s}\rangle$ for a given primary ϕ_j is called *the Verma module of the highest weight state* $|\phi_j\rangle$. The non-negative integer $n \equiv \sum_k n_k$ is called the *level* of the state in its Verma module. States of positive level in the Verma module of the primary $|\phi_j\rangle$ are called the *descendants* of $|\phi_j\rangle$.

³¹ Properly speaking they are *lowest* weight states; mathematicians use the weight $-h$ so for them the primary states are highest weight vectors, and their language is used abusively by most physicists.

As a Virasoro representation, the Verma module of $|\phi_j\rangle$ depends, up to isomorphism, only on the values of c and h_j . We write $V(c, h_j)$ for this “abstract” module.

As we shall see momentarily, in a “good”³² CFT L_0 is a *non-negative* operator. Acting on a L_0 -eigenstate with the operators L_n ($n > 0$) *decreases* the eigenvalue of L_0 by $n \geq 1$. Then, after applying finitely many L_n ’s, we reach the highest weight state with $L_n|\psi\rangle = 0$ for all $n > 0$, i.e. *in a “good” CFT all non-primary states are the descendants of some primary state*. By the state-operator isomorphism, all non-primary operators are the descendants of some primary operator. This has the important consequence that the n -point functions of *all* local operators are determined once we know the correlators of primaries.

We use a basis $\{|\phi_j\rangle\}$ of primary states which is orthonormal

$$\langle\phi_j^\dagger|\phi_k\rangle = \delta_{jk}. \quad (2.143)$$

States belonging to Verma modules of different elements of $\{|\phi_j\rangle\}$ are then orthogonal, as are states in the same Verma module having different levels.

The operators corresponding to the states of the Verma module generated by the primary ϕ_j constitute the *conformal family* $[\phi_j]$ of ϕ_j

$$[\phi_j] = \{L_{-n_1} \cdots L_{-n_s} \cdot \phi_j(z)\}. \quad (2.144)$$

Note that the energy–momentum tensor is a descendant of the identity, $T = L_{-2} \cdot 1$.

Partitions Let $P(n)$ be the number of operators of the form

$$L_{-n_1} L_{-n_2} \cdots L_{-n_s}, \quad n_1 \geq n_2 \geq \cdots \geq n_s > 0 \quad (2.145)$$

with a given level $n \equiv \sum_k n_k$. $P(n)$ is the *number of partitions of n* , i.e. the number of ways we can write n as a sum of positive integers. For the values of $P(n)$, see sequence A000041 in [30]; for mathematical properties of partitions, see [31, 32]. The asymptotics of $P(n)$ for large n is given by the *Hardy–Ramanujan formula* [31]

$$P(n) \approx \frac{1}{4\sqrt{3}n} \exp\left[\pi\sqrt{2n/3} + O(\log(n^{-1/4} \log n))\right]. \quad (2.146)$$

We do not prove this formula here, since it will be an obvious corollary of the explicit expression of the partition function for a free massless scalar, Eq. (4.60) in view of the Cardy formula (2.174).

Exercise 2.5 Show that the generating function of $P(n)$ is

$$\sum_{n \geq 0} P(n) q^n = \prod_{n=1}^{\infty} (1 - q^n)^{-1}. \quad (2.147)$$

³² All unitary CFT are “good” in the present sense.

Characters: Singular States

The character $\chi_j(\tau)$ of the irreducible representation $[\phi_j]$ of the Virasoro algebra whose highest weight vector is the primary state $|\phi_j\rangle$ is the generating function for the degeneracy $d_j(n)$ of level- n states in the irreducible module $[\phi_j]$:

$$\chi_j(\tau) \stackrel{\text{def}}{=} \text{Tr}_{[\phi_j]} \left[q^{L_0 - c/24} \right] = q^{h_j - c/24} \sum_{n=0}^{\infty} d_j(n) q^n, \quad \text{with } q \equiv e^{2\pi i \tau}. \quad (2.148)$$

In a *generic* Verma module the states (2.141) are linearly independent, and hence

$$d_j(n) = P(n) \quad (\text{generic Verma module}). \quad (2.149)$$

However it may happen that, for some special values of c and h_j , there are states $|\chi\rangle \in V(c, h_j)$ which are themselves primary, that is, annihilated by all L_n with $n > 0$. Such states, which are both primary and descendant, are called *singular* or *null*. A singular state $|\chi\rangle$ is orthogonal to all states in $V(c, h_j)$ including itself,³³ That is, it has zero “norm”, $\langle \chi^\dagger | \chi \rangle = 0$.

A singular state $|\chi\rangle$ is the highest weight state of its own Verma module V_χ ; therefore a Verma module containing singular states is *not* an irreducible Virasoro representation. To get an irreducible representation of the Virasoro algebra we need to take the quotient with respect to all non-trivial proper subrepresentations, that is,

$$V(c, h_j)_{\text{irred.}} = V(c, h_j) \bigg/ \sum_{\substack{|\chi\rangle \in V(c, h_j) \\ \text{singular}}} V_\chi, \quad (2.150)$$

i.e. we must declare two states in $V(c, h)$ to be equivalent if they differ by a null state

$$|\psi\rangle \sim |\psi\rangle + |\text{null}\rangle. \quad (2.151)$$

Then, for all characters $\chi_j(\tau)$ of irreducible Virasoro representations,

$$d_j(n) \leq P(n). \quad (2.152)$$

2.3.5 Unitarity

Unitarity of a Virasoro representation V means that the invariant Hermitian product (2.134) is positive-definite on V . In particular this requires ($|\phi_j\rangle$ primary, $n > 0$)

³³ **Pf.** Being a descendant $|\chi\rangle = L_{-n}|\psi\rangle$ for some $n > 0$ and $|\psi\rangle$. Then

$$\langle \chi^\dagger | \chi \rangle = \langle \chi^\dagger | L_{-n} | \psi \rangle = \left(\langle \psi^\dagger | L_n | \chi \rangle \right)^* = 0 \quad \text{since } |\chi\rangle \text{ is also primary.}$$

$$0 \leq \|L_{-n}|\phi_j\rangle\|^2 \equiv \langle\phi_j^\dagger|L_n L_{-n}|\phi_j\rangle = \left[2n h_j + \frac{c}{12}(n^3 - n)\right] \langle\phi_j^\dagger|\phi_j\rangle \quad (2.153)$$

with equality iff $L_{-n}|\phi_j\rangle = 0$. Barring this singular case, taking $n \gg 1$ we see that in a unitary CFT we must have $c \geq 0$, while setting $n = 1$ we see that $h_j \geq 0$ with $h_j = 0$ if and only if $L_{-1}|\phi_j\rangle \equiv \partial\phi_j(0)|0\rangle = 0$. This shows that in a unitary CFT the identity is the only primary operator with $h = 0$ (cf. footnote 29).

A *non-trivial* unitary CFT should have $c > 0$. Indeed

$$\langle T^\dagger | T \rangle = \langle 0 | L_2 L_{-2} | 0 \rangle = \langle 0 | [L_2, L_{-2}] | 0 \rangle = \frac{c}{2}, \quad (2.154)$$

so $c = 0$ means $|T\rangle = 0$, and then, by the state-operator isomorphism, $T(z) \equiv 0$ as an operator. As a consequence all primaries have $h = 0$, hence the only operator in the theory is the identity.

Note 2.4 There are many interesting non-trivial *non-unitary* CFT with $c \leq 0$. For example, the (gauge-fixed) world-sheet theory of the bosonic string³⁴ has $c = 0$, but it is non-unitary because of the wrong-statistics Faddeev–Popov ghosts b, c .

When $c \geq 1$ unitarity gives no other constraint on h_j , i.e. the Verma modules $V(c, h_j)$ have positive-definite inner product for $c \geq 1, h_j \geq 0$. For $c < 1$ the story is much subtler. In the range $0 < c < 1$ unitary CFTs exist only for very special values of c (the so-called *unitary series*) [1, 3]

$$c = 1 - \frac{6}{m(m+1)}, \quad m = 3, 4, 5, \dots \quad (2.155)$$

and for a given m there is only a finite set of allowed weights h

$$h_{p,q} = \frac{[(m+1)p - mq]^2 - 1}{4m(m+1)}, \quad (2.156)$$

$$p = 1, 2, \dots, m-1, \quad q = 1, 2, \dots, p.$$

Unitary CFT with $c < 1$ are called (*Virasoro*) *minimal models*. They are fully understood (all correlation functions are known). For details see, for example, [3].

Exercise 2.6 Show that in a *unitary* CFT $L_0|\psi\rangle = |\psi\rangle$ implies $\psi(z)$ is primary.

2.3.6 General Chiral Algebras in 2d CFT

A unitary 2d CFT may have several chiral operators $J(z)$ of weights $(h, 0)$ besides the energy–momentum tensor $T(z)$. They all satisfy $\bar{\partial}J(z) = 0$, i.e. they are conserved

³⁴ Indeed the Weyl anomaly is measured by c , Eq. (2.103), while in the string case the Weyl anomaly cancels between the “matter” sector and the ghosts.

chiral currents of spin h . Unitarity implies that $J(z)$ is bosonic for h integral and fermionic for h half-integral. If $\epsilon(z)$ is a holomorphic section of K^{1-h} , $\epsilon(z)J(z)$ is a holomorphic 1-form, and may be integrated along closed contours to produce a conserved charge Q_ϵ such that, for all local operators $\Phi(z)$,³⁵

$$\delta_\epsilon \Phi(z) \equiv [Q_\epsilon, \Phi(z)] = \oint_{C_z} \frac{dw}{2\pi i} \epsilon(w) J(w) \Phi(z). \quad (2.157)$$

We get a Ward identity for the correlation functions from each conserved current $\epsilon(z)J(z)$ by deforming the contour of integration in its homology class just as we did in Sect. 2.1 for the current $T(z)v(z)$ associated with a holomorphic vector field $v(z)$.

The set of all chiral currents generates a (left-moving) *chiral algebra* $\mathcal{A}$ containing the Virasoro algebra and carrying its adjoint representation. Besides the Virasoro algebra itself, several other chiral algebras are important for 2d QFT and string theory: some of them will be discussed in later sections of this chapter.

On the right side we have a second (right-moving) chiral algebra $\tilde{\mathcal{A}}$, not necessarily isomorphic to $\mathcal{A}$. The Hilbert space of the theory splits into a direct sum of irreducible representations of the chiral algebras

$$\mathbf{H} = \bigoplus_{a, \tilde{b}} (\mathcal{H}_a \otimes \tilde{\mathcal{H}}_{\tilde{b}})^{\oplus m(a, \tilde{b})} \quad (2.158)$$

where $\mathcal{H}_a$ (resp. $\tilde{\mathcal{H}}_{\tilde{b}}$) is an irreducible representation of $\mathcal{A}$ (resp. $\tilde{\mathcal{A}}$), the sum is over all pairs of isoclasses of irreducible representations, and the non-negative integer $m(a, \tilde{b})$ is the multiplicity of the given representation pair.

By definition, a *Rational Conformal Field Theory* (RCFT) is a CFT such that there are *finitely many* irreducible representations of its left- and right-chiral algebras $\mathcal{A}$ and $\tilde{\mathcal{A}}$, so that the direct sum in (2.158) is finite [3] (for a review, see [33]). For example, the *minimal models* alluded to at the end of the previous subsection are RCFT whose chiral algebra is the Virasoro algebra itself.

Just as in Eq. (2.148) for the Virasoro algebra, we may define characters χ_a for all irreducible representations $\mathcal{H}_a$ of $\mathcal{A}$

$$\chi_a(q) = \text{Tr}_{\mathcal{H}_a} \left[q^{L_0 - c/24} \right]. \quad (2.159)$$

More generally, we introduce chemical potentials z_i for a maximal set of commuting bosonic charges J_0^i associated with the currents in $\mathcal{A}$ with $[L_0, J_0^i] = 0$

$$\chi_a(q, z_i) = \text{Tr}_{\mathcal{H}_a} \left[q^{L_0 - c/24} \exp \left(\sum_i z_i J_0^i \right) \right]. \quad (2.160)$$

³⁵ The contour C_z encircles the point z .

BOX 2.4 - Complex moduli of tori and modular invariance

By uniformization (BOX 1.9), all genus one Riemann surface has the form $E = \mathbb{C}/\Lambda$ where $\Lambda \subset \mathbb{C}$ is a *lattice*, i.e. a subgroup of the form $\{ma + nb\}_{m,n \in \mathbb{Z}}$ with a, b complex numbers linearly independent over $\mathbb{R}$. Without loss we assume $\tau \equiv a/b$ to have a positive imaginary part. Rescaling the coordinate z of $\mathbb{C}$ we take $b = 1$, i.e. E is $\mathbb{C}$ modulo the equivalence relation

$$z \sim z + m\tau + n, \quad m, n \in \mathbb{Z}, \quad \text{that is } \Lambda \equiv \tau\mathbb{Z} \oplus \mathbb{Z}.$$

τ takes value in the upper half-plane $\mathcal{H} = \{z \in \mathbb{C} : \text{Im } z > 0\}$. One has

$$H_1(E, \mathbb{Z}) \equiv \pi_1(E)^{\text{ab}} \equiv \pi_1(\mathbb{C}/\Lambda)^{\text{ab}} \simeq \Lambda^{\text{ab}} \equiv \Lambda = \tau\mathbb{Z} \oplus \mathbb{Z}$$

where G^{ab} stands for the abelianization $G/[G, G]$ of the group G . The isomorphism is given by the integral of the closed 1-form dz along the cycle. As basis of $H_1(E, \mathbb{Z})$ we may choose the closed loops $A = \{z = t, 0 \leq t \leq 1\}$ and $B = \{z = t\tau, 0 \leq t \leq 1\}$ with periods

$$\int_A dz = 1, \quad \int_B dz = \tau.$$

$\{A, B\}$ form a symplectic basis of $H_1(E, \mathbb{Z})$, i.e. the cycles have the intersection form

$$A \cdot A = B \cdot B = 0, \quad A \cdot B = -B \cdot A = 1.$$

A point $\tau \in \mathcal{H}$ then parametrizes a *marked elliptic curve*, i.e. a pair $(E, \{A, B\})$ where E is an elliptic curve (a torus with a specified complex structure) and $\{A, B\}$ a choice of symplectic basis for $H_1(E, \mathbb{Z})$. On the set of all symplectic bases the *modular group*

$$SL(2, \mathbb{Z}) \equiv Sp(2, \mathbb{Z})$$

acts transitively as

$$\begin{bmatrix} B' \\ A' \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} B \\ A \end{bmatrix}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}),$$

so that

$$\tau' = \frac{\int_{B'} dz}{\int_{A'} dz} = \frac{a\tau + b}{c\tau + d} \quad \text{modular transformation.}$$

The quotient group acting effectively on $\mathcal{H}$ is $PSL(2, \mathbb{Z}) \equiv SL(2, \mathbb{Z})/\{\pm 1\}$. The moduli space of genus one Riemann surfaces is then

$$\mathcal{M}_1 = \mathcal{M}_{1,1} = \mathcal{H}/PSL(2, \mathbb{Z}).$$

A fundamental domain $\mathcal{F}$ of the modular group $PSL(2, \mathbb{Z})$ is a region in $\mathcal{H}$ such that all point $x \in \mathcal{M}_1$ has a representative modulo $PSL(2, \mathbb{Z})$ in $\mathcal{F}$, and two points $\tau_1, \tau_2 \in \mathcal{F}$ are not in the same $PSL(2, \mathbb{Z})$ -orbit unless they both lay on the boundary $\partial\mathcal{F}$ of $\mathcal{F}$. The domain

$$\mathcal{F} = \{\tau \in \mathcal{H} : -\frac{1}{2} \leq \text{Re } \tau \leq \frac{1}{2}, \quad |\tau| \geq 1\} \subset \mathcal{H}$$

is a fundamental domain for the modular group $PSL(2, \mathbb{Z})$.

For more details on the modular group, see Sect. 5.1.

2.3.7 Partition Function and Modular Invariance

We use the standard notations

$$q = e^{2\pi i \tau}, \quad \bar{q} = e^{-2\pi i \bar{\tau}}, \quad (2.161)$$

where $\tau = \tau_1 + i\tau_2$ is a point in the *upper half-plane* $\mathcal{H}$:

$$\mathcal{H} \stackrel{\text{def}}{=} \left\{ \tau \in \mathbb{C} : \tau_2 \equiv \text{Im } \tau > 0 \right\}. \quad (2.162)$$

The *partition function* of a 2d CFT is the trace in the Hilbert space $\mathbf{H}$ of the operator

$$q^{L_0 - c/24} \bar{q}^{\tilde{L}_0 - c/24} \equiv \exp \left[-2\pi \tau_2 H_{\text{cyl}} + 2\pi i \tau_1 P_{\text{cyl}} \right], \quad (2.163)$$

where we used the identifications

$$H_{\text{cyl}} = L_0 + \tilde{L}_0 - \frac{c}{12}, \quad P_{\text{cyl}} = L_0 - \tilde{L}_0. \quad (2.164)$$

Hence Eq. (2.163) corresponds to an inverse temperature $\beta = 2\pi \tau_2$, while the fugacity of the conserved charge P_{cyl} (the momentum along the spatial circle) is $\lambda = e^{2\pi i \tau_1}$.

From Eqs. (2.158), (2.159) we get

$$Z(\tau, \bar{\tau}) = \sum_{a, \tilde{b}} m(a, \tilde{b}) \chi_a(\tau) \tilde{\chi}_{\tilde{b}}(\bar{\tau}). \quad (2.165)$$

In a RCFT the sum is finite. The coefficients $m(a, \tilde{b})$ are non-negative integers which for “good” CFTs are further restricted by *modular invariance*.

Modular Invariance

The partition function $Z(\tau, \bar{\tau})$ is also given by the path integral of the 2d CFT (without insertions) on an infinite cylinder $\mathbb{R} \times S^1$ where we periodically identify the Euclidean coordinates $(t, \theta) \sim (t + \beta, \theta + 2\pi \tau_1)$. Geometrically, this is equivalent to the path integral on a 2-torus, i.e. on $S^1 \times S^1$. The partition function of a 2d CFT depends on the 2d metric only through its conformal class, that is, its underlying *complex structure*. As explained in **BOX 2.4**, the modulus $\tau \in \mathcal{H}$ parametrizes complex structures of *marked* tori. In a *bona fide* CFT the partition function cannot depend on a choice of marking. Indeed two distinct markings are related by a “large” diffeomorphism,³⁶ so $Z(\tau, \bar{\tau})$ may depend on the marking only if the model is not invariant under “large” diffeomorphisms, that is, if the CFT suffers from 2d global gravitational anomalies. This cannot happen in a “good” local theory.³⁷ Then we must have

³⁶ A diffeomorphism is “large” if it is not homotopic to the identity.

³⁷ There are plenty of non-“good” (interesting) theories with **Diff** anomalies (both local and global).

$$Z\left(\frac{a\tau + b}{c\tau + d}, \frac{a\bar{\tau} + b}{c\bar{\tau} + d}\right) = Z(\tau, \bar{\tau}), \quad \begin{cases} \text{for } a, b, c, d \in \mathbb{Z} \\ \text{with } ad - bc = 1. \end{cases} \quad (2.166)$$

This condition is called *modular invariance*. It suffices to check the two special cases

$$Z(\tau, \bar{\tau}) = Z(-1/\tau, -1/\bar{\tau}) = Z(\tau + 1, \bar{\tau} + 1), \quad (2.167)$$

since the full modular group $SL(2, \mathbb{Z})$ is generated by $S: \tau \mapsto -1/\tau$ and $T: \tau \mapsto \tau + 1$. Modular invariance and its consequences for string theory will be considered in great detail in Chap. 5 from various viewpoints.

Modular invariance is a formidable constraint on the weights $(h_a, \tilde{h}_b)$ of the primary fields and their multiplicities $m(a, \tilde{b}) \in \mathbb{N}$ (cf. Eq. (2.165)). For instance

Theorem 2.2 (Cardy [34]) *A modular invariant, unitary, 2d CFT with Virasoro central charge $c \geq 1$ has infinitely many primary operators.*

In the opposite situation, $c < 1$, the CFT (if unitary) should be a minimal model (2.155) which has only finitely many primaries; see, for example, [3].

To prove the **Theorem** we first establish the *Cardy asymptotic formula*, a result of independent interest.

The Cardy Asymptotic Formula

To make the physical meaning more transparent we change notations: we write L for the length of the circle on which we define the theory (usually normalized to $L = 2\pi$), β for the period of the Euclidean time (i.e. the inverse temperature), and set the momentum chemical potential τ_1 to zero. Then $\tau = i\beta/L$, and the cylinder Hamiltonian becomes $H(L) = 2\pi H_{\text{cyl}}/L$. The first equality (2.167) gives a relation between the partition functions at high and low temperatures

$$\text{Tr}\left[e^{-\beta H(L)}\right] \equiv \text{Tr}\left[e^{-2\pi(\beta/L)H_{\text{cyl}}}\right] = \text{Tr}\left[e^{-2\pi(L/\beta)H_{\text{cyl}}}\right]. \quad (2.168)$$

We take $\beta \rightarrow 0$ in this equality. In the RHS the temperature $T \equiv \beta/(2\pi L)$ goes to zero, and only the vacuum contributes. The eigenvalue of H_{cyl} acting on the vacuum is $E_0 = -(c + \tilde{c})/24$, so

$$\text{Tr}\left[e^{-\beta H(L)}\right] \approx \exp\left[\frac{\pi(c + \tilde{c})L}{12\beta}\right] \quad \text{as } \beta \rightarrow 0. \quad (2.169)$$

On the other hand, the LHS of (2.168) can be written in terms of the density of states $n(E, L)$ with energy $E = E' + 2\pi E_0/L$ in a circle of length L as

$$e^{2\pi\beta E_0/L} \text{Tr}\left[e^{-\beta H(L)}\right] = \int_0^\infty dE' n(E, L) e^{-\beta E'}, \quad (2.170)$$

that is, inverting the Laplace transform,

$$n(E, L) = \int_{a-i\infty}^{a+i\infty} \frac{d\beta}{2\pi i} e^{\beta E} \text{Tr} \left[e^{-\beta H(L)} \right]. \quad (2.171)$$

As $E \rightarrow \infty$ the integral localizes at $\beta \approx 0$ and from Eq. (2.169), we get

$$n(E, L) \approx \int_{a-i\infty}^{a+i\infty} \frac{d\beta}{2\pi i} e^{\beta E} \exp \left[\frac{\pi(c + \tilde{c})L}{12\beta} \right]. \quad (2.172)$$

In the limit $E \rightarrow \infty$ the integral may be computed by saddle-point methods. The saddle point is at

$$\beta^2 = \frac{\pi(c + \tilde{c})L}{12E}, \quad (2.173)$$

and the *Cardy formula* for the asymptotic density of states for large energy E is

$$n(E, L) \approx \frac{1}{2\sqrt{\pi}} \left(\frac{\pi(c + \tilde{c})L}{12} \right)^{1/4} E^{3/4} \cdot \exp \left[\sqrt{\frac{(c + \tilde{c})\pi}{3}} L E \right]. \quad (2.174)$$

Proof (of Cardy theorem) We return to the standard normalization $L = 2\pi$, so

$$n(E) \approx \text{const.} E^{3/4} \exp \left[\pi \sqrt{2(c + \tilde{c})E/3} \right]. \quad (2.175)$$

Consider the partition function written as the sum (2.165) of Virasoro characters. The $(j, \tilde{i})$ term in the sum contributes $d_j(E/2) d_{\tilde{i}}(E/2)$ to $n(E)$, where $d_j(n)$ are the multiplicities in (2.148). Using the bound (2.152) and the Hardy–Ramanujan formula (2.146), we get as $E \rightarrow \infty$

$$d_j(E/2) d_{\tilde{i}}(E/2) \leq \text{const.} \frac{1}{E} \exp \left[2\pi \sqrt{E/3} \right]. \quad (2.176)$$

The exponential factor is equal to the one in the Cardy formula (2.175) for $c \equiv \tilde{c} = 1$. Thus when (2.165) is a finite sum, the function $n(E)$ grows with E at most as a constant times $\exp \left[2\pi \sqrt{E/3} \right] / E$ which is inconsistent with (2.175) for $c \equiv \tilde{c} \geq 1$. $\square$

Note 2.5 The proof shows that, whenever $c < 1$, the Verma modules of the primaries contains a lot of null states.

2.3.8 More on Correlation Functions. Normal Order

Correlation functions of quasi-primary operators on the sphere S^2 are invariant under the global $SL(2, \mathbb{C})$. This leads to the following three identities:

$$\sum_i \partial_{z_i} \langle \phi_1(z_1) \cdots \phi_n(z_n) \rangle = 0 \quad (2.177)$$

$$\sum_i (z_i \partial_{z_i} + h_i) \langle \phi_1(z_1) \cdots \phi_n(z_n) \rangle = 0 \quad (2.178)$$

$$\sum_i (z_i^2 \partial_{z_i} + 2z_i h_i) \langle \phi_1(z_1) \cdots \phi_n(z_n) \rangle = 0. \quad (2.179)$$

Equation (2.177) is translational invariance that says that the correlation function depends only on $z_i - z_j$. One-point functions are then constant and Eq. (2.178) becomes

$$h_i \langle \phi_i(0) \rangle = 0 \quad (2.180)$$

so that *the only quasi-primary operator with a non-zero v.e.v. on S^2 is the identity*.

For the 2-point functions we get

$$\langle \phi_i(z) \phi_j(w) \rangle = \begin{cases} G_{ij}(z-w)^{-2h_i} & \text{for } h_i = h_j \\ 0 & \text{otherwise} \end{cases} \quad (2.181)$$

for some non-degenerate quadratic form G_{ij} . It is convenient to normalize our basis of primary fields to be orthonormal

$$\langle \phi_i^\dagger(z) \phi_j(w) \rangle = \frac{\delta_{ij}}{(z-w)^{2h_j}}. \quad (2.182)$$

The 3-point function has the form ($z_{ij} \equiv z_i - z_j$)

$$\langle \phi_i(z_1) \phi_j(z_2) \phi_k(z_3) \rangle = \frac{C_{ijk}}{z_{12}^{h_i+h_j-h_k} z_{13}^{h_i+h_k-h_j} z_{23}^{h_j+h_k-h_i}} \quad (2.183)$$

for some numerical constants C_{ijk} . From the 3-point functions we can compute the expectation values between asymptotic states

$$\langle \phi_i \text{ out} | \phi_j(z) | \phi_k \text{ in} \rangle = \frac{C_{ijk}}{z^{h_j+h_k-h_i}}. \quad (2.184)$$

The functional form of n -point functions with $n \geq 4$ is not determined by symmetry considerations alone. Indeed, we saw in Chap. 1 that in CFT an n -point correlator on the sphere is determined by conformal invariance up to a non-trivial function on the complex manifold³⁸ $\mathcal{M}_{0,n}$. Now, while $\mathcal{M}_{0,3}$ is a single point, a function on $\mathcal{M}_{0,3}$ is just a complex number that we call C_{ijk} , for $n \geq 4$ the space $\mathcal{M}_{0,n}$ has *positive dimension*

$$\dim_{\mathbb{C}} \mathcal{M}_{0,n} = n - 3 > 0 \quad \text{for } n \geq 4, \quad (2.185)$$

and the n -point correlation depends on a function of $n - 3$ complex arguments which is not restricted by symmetry.

³⁸ Recall from Chap. 1 that $\mathcal{M}_{g,n}$ stands for the moduli space of complex structures on a compact Riemann surface of genus g with n punctures.

Example: The 4-point function

As a local coordinate on $\mathcal{M}_{0,4}$ we may use the *cross-ratio*:

$$\frac{z_{12} z_{24}}{z_{13} z_{24}}, \quad \text{where } z_{ij} \equiv z_i - z_j, \quad (2.186)$$

and write the 4-point function in the form

$$\langle \phi_1(z_1) \phi_2(z_2) \phi_3(z_3) \phi_4(z_4) \rangle = \frac{z_{13}^{h_3+h_4} z_{24}^{h_1+h_3}}{z_{12}^{h_1+h_2} z_{23}^{h_2+h_3} z_{34}^{h_3+h_4} z_{14}^{h_1+h_2}} f\left(\frac{z_{12} z_{24}}{z_{13} z_{24}}\right) \quad (2.187)$$

where the function f is not restricted by conformal symmetry.

A 2d CFT is fully determined by the following set of data: the central charge c , the weights $\{h_j\}$ of all the primary operators, and the 3-point coefficients C_{ijk} between primary operators. Not all set of conformal data $\{c, h_j, C_{ijk}\}$ yield a sound unitary CFT. The *conformal bootstrap* [22–25] is a technique which aims to determine the sets of conformal data which *do define* a meaningful CFT.

Normal Order We define the *conformal normal order*

$$: \phi_i(z) \phi_j(z) : = \sum_n \frac{(: \phi_i \phi_j :)_n}{z^{n+h_i+h_j}} \quad (2.188)$$

of the product of two *mutually local* operators $\phi_i(z)$ and $\phi_j(z)$, i.e. two operators such that the only singularities in $\phi_i(z)\phi_j(w)$ are poles. Their conformal normal product is their OPE with the singular part subtracted. For chiral operators this reduces to

$$\begin{aligned} : \phi_i(z) \phi_j(z) : &\stackrel{\text{def}}{=} \lim_{w \rightarrow z} \left(\phi_i(w) \phi_j(z) - \text{polar part of OPE} \right) \equiv \\ &\equiv \oint \frac{dw}{2\pi i} \frac{\phi_i(w) \phi_j(z)}{z - w}. \end{aligned} \quad (2.189)$$

Plugging in the RHS the expansion (2.90), we get the modes of the normal product

$$(: \phi_i \phi_j :)_m = \sum_{n \leq -h_i} \phi_{i,n} \phi_{j,m-n} + \sum_{n > -h_i} \phi_{j,m-n} \phi_{i,n} \quad (2.190)$$

and, using Eq. (2.130),

$$\langle 0 | : \phi_i \phi_j : | 0 \rangle = 0, \quad (2.191)$$

that is, the conformal normal order $: \dots :$ is defined with respect to the $SL(2, \mathbb{C})$ -invariant vacuum $|0\rangle$. Sometimes one uses normal orders with respect to other states (e.g. sea levels; see Sect. 2.5.3); to avoid confusion we shall reserve the notation $: \dots :$ for the conformal normal order (2.189). More in general we have

$$\phi_i(z)\phi_j(w) = [\phi_i(z)\phi_j(w)]_{\text{sing}} + : \phi_i(z)\phi_j(w) :, \quad (2.192)$$

where the first term in the RHS is the singular part of the OPE.

2.4 Example: The 2d Free Massless Scalar

A simple CFT is the free massless scalar in 2d with action

$$S = \frac{1}{2\pi\alpha'} \int d^2z \, \partial X \bar{\partial} X, \quad (2.193)$$

where the 2d scalar X may be *non-compact*, i.e. valued in $\mathbb{R}$, or *compact*, that is, valued in a circle of radius R , i.e. periodically identified

$$X \sim X + 2\pi R. \quad (2.194)$$

The CFT (2.193) is a basic ingredient of string theory: e.g. the world-sheet theory of the bosonic string in the light-cone gauge consists of 24 copies of the non-compact scalar. In this section we focus on the *non-compact* CFT, deferring the study of compact scalars and related stringy phenomena to later sections and Chap. 6.

Basic Formulae

We consider the free scalar in radial quantization. The equations of motion are

$$\partial \bar{\partial} X(z, \bar{z}) = 0 \quad (2.195)$$

whose a general solution is

$$X(z, \bar{z}) = X_L(z) + X_R(\bar{z}) \quad \begin{array}{l} \text{decomposition in} \\ \text{left- and right-movers,} \end{array} \quad (2.196)$$

with $X_L(z)$, $X_R(\bar{z})$ arbitrary holomorphic functions. The mode expansion then reads

$$X(z, \bar{z}) = q - i \frac{\alpha'}{2} p \log |z|^2 + i \sqrt{\frac{\alpha'}{2}} \sum_{n \neq 0} \left(\frac{\alpha_n}{n z^n} + \frac{\tilde{\alpha}_n}{n \bar{z}^n} \right). \quad (2.197)$$

The 2-point function on the sphere is

$$\langle X(z, \bar{z}) X(w, \bar{w}) \rangle = -\frac{\alpha'}{2} \log |(z - w)/\Lambda|^2, \quad (2.198)$$

where the scale Λ may be seen as an IR regulator. Conformal invariance requires Λ to drop out from all correlation functions of local operators. Equation (2.198) then

BOX 2.5 - No spontaneous symmetry breaking in 2d

The Coleman theorem [35] (a.k.a. Mermin–Wagner [36]) states that in 2d a continuous bosonic symmetry cannot be broken spontaneously. In fact, if the symmetry associated with the Noether current J_μ was spontaneously broken, by the Goldstone theorem [37] we must have a massless scalar state $|\pi\rangle$ such that $\langle 0|J_\mu|\pi\rangle = ip_\mu F_\pi \neq 0$ at momentum p_μ , while the variation $\delta\phi = [Q, \phi]$ of the order parameter ϕ satisfies $\langle \pi|\delta\phi|0\rangle = \sigma \neq 0$. In this case the correlation function $\langle 0|J_\mu(z)\delta\phi(w)|0\rangle$ would have a IR singularity so severe that it would not be a distribution. (Showing this fact is left as an useful exercise.) In particular, in 2d a free scalar X cannot be a local operator: indeed, if it was, its 1-point function $\langle X \rangle$ would be defined and this fact would break the *unbreakable* shift symmetry $X \rightarrow X + \text{const.}$ Instead the exponentials e^{ipX} are well-defined local operators with $\langle e^{ipX} \rangle = 0$ if $p \neq 0$, as a consequence of the fact that the constant mode of X fluctuates in a wild way. In the application to string theory this result corresponds to momentum conservation in physical (target) spacetime. In other words,

The Coleman theorem guarantees that in string theory the *spacetime* Diff^+ symmetry is **not** spontaneously broken, a crucial requirement for a theory containing Einstein gravity.

says that $X(z, \bar{z})$ is **not** a well-defined local operator: its n -point functions are not tempered distributions, [35] a fact that implies Coleman’s theorem (see BOX 2.5).

Equation (2.198) is a solution to the PDE in the plane

$$\partial_z \bar{\partial}_{\bar{z}} \langle X(z, \bar{z}) X(w, \bar{w}) \rangle = -\pi \alpha' \delta^{(2)}(z - w). \quad (2.199)$$

From Eq.(2.196) we see that the operators $\partial X(z)$ and $\bar{\partial} X \equiv \bar{\partial} \tilde{X}(\bar{z})$ are chiral fields with mode expansions

$$\partial X(z) = \sum_{n \in \mathbb{Z}} \frac{\alpha_n}{z^{n+1}}, \quad \bar{\partial} X(\bar{z}) = \sum_{n \in \mathbb{Z}} \frac{\tilde{\alpha}_n}{\bar{z}^{n+1}} \quad (2.200)$$

where for non-compact scalars we adopt the convention

$$\alpha_0 = \tilde{\alpha}_0 = \sqrt{\frac{\alpha'}{2}} p \quad (2.201)$$

with p the target-space momentum (i.e. the Noether charge associated with the shift symmetry $X \rightarrow X + \text{const.}$). Their 2-point functions are

$$\langle \partial X(z) \partial X(w) \rangle = -\frac{\alpha'}{2} \frac{1}{(z - w)^2}, \quad \langle \bar{\partial} \tilde{X}(\bar{z}) \bar{\partial} \tilde{X}(\bar{w}) \rangle = -\frac{\alpha'}{2} \frac{1}{(\bar{z} - \bar{w})^2}, \quad (2.202)$$

corresponding to the OPE

$$\begin{aligned}
 \partial X(z) \partial X(w) &= -\frac{\alpha'}{2} \frac{1}{(z-w)^2} + : \partial X(z) \partial X(w) : = \\
 &= -\frac{\alpha'}{2} \frac{1}{(z-w)^2} + : \partial X(w) \partial X(w) : + \\
 &\quad + \frac{1}{2}(z-w) \partial (: \partial X(w) \partial X(w) :) + O((z-w)^2).
 \end{aligned} \tag{2.203}$$

In terms of modes this OPE becomes the canonical commutator algebra

$$[\alpha_m, \alpha_n] = m \delta_{m+n,0} \tag{2.204}$$

as already discussed in Chap. 1 in the context of light-cone quantization.

Exercise 2.7 Check the equivalence of Eqs. (2.203) and (2.204).

We shall see momentarily that $\partial X(z)$ is a conformal primary with $(h, \tilde{h}) = (1, 0)$, i.e. $\partial X(z)$ transforms as a holomorphic differential (section of the canonical bundle K).

The left-moving energy–momentum tensor is

$$T(z) = -\frac{1}{\alpha'} : \partial X(z) \partial X(z) :. \tag{2.205}$$

A similar formula holds for its right-moving counterpart $\tilde{T}(\bar{z})$. The mode expansion (2.200), yields the Virasoro generators

$$L_n \equiv \oint \frac{dz}{2\pi i z} z^{n+2} T(z) \tag{2.206}$$

in terms of the modes of X

$$L_n = \frac{1}{2} \sum_{m \in \mathbb{Z}} : \alpha_{n-m} \alpha_m : \equiv \frac{1}{2} \sum_{n \leq -1} \alpha_n \alpha_{m-n} + \frac{1}{2} \sum_{n > -1} \alpha_{m-n} \alpha_n. \tag{2.207}$$

In particular,

$$L_0 = \frac{\alpha'}{4} p^2 + \sum_{n \geq 1} \alpha_{-n} \alpha_n. \tag{2.208}$$

We compute the OPE of two $T(z)$: using Wick theorem and Eq. (2.203)

$$\begin{aligned}
T(z) T(w) &= \frac{2}{(\alpha')^2} \left(\partial X(z) \partial X(w) \right)^2 = \\
&= \frac{1}{2} \left(-\frac{1}{(z-w)^2} - 2 T(w) - (z-w) \partial T(w) + \dots \right)^2 = \quad (2.209) \\
&= \frac{1}{2(z-w)^4} + \frac{2 T(w)}{(z-w)^2} + \frac{\partial T(w)}{(z-w)} + \text{regular}
\end{aligned}$$

which is the Virasoro OPE with central charge $c = 1$. This result confirms that the non-compact scalar (2.193) is indeed a CFT with $c = 1$.

We compute the OPE of $T(z)$ with the operator $\partial X(w)$

$$\begin{aligned}
T(z) \partial X(w) &= -\frac{2}{(\alpha')^2} \partial X(z) [\partial X(z) \partial X(w)]_{\text{sing.}} + \text{reg.} = \\
&= -\frac{2}{(\alpha')^2} \left(\partial X(w) + (z-w) \partial^2 X(w) + \dots \right) \left(-\frac{\alpha'}{2} \frac{1}{(z-w)^2} \right) + \text{reg.} \quad (2.210) \\
&= \frac{\partial X(w)}{(z-w)^2} + \frac{\partial(\partial X(w))}{(z-w)} + \text{regular}.
\end{aligned}$$

This OPE justifies our claim that $\partial X(z)$ is a *conformal primary* with $(h, \tilde{h}) = (1, 0)$.

Relation to Abelian Currents

A conformal primary $J(z)$ with $(h, \tilde{h}) = (1, 0)$ is a *conserved chiral current*.³⁹ Indeed

$$\bar{\partial} J(z) \equiv [\tilde{L}_{-1}, J(z)] = 0. \quad (2.211)$$

The corresponding Abelian charge is

$$Q = \oint_C \frac{dz}{2\pi i} J(z), \quad [Q, \phi(w)] = \oint_{C_w} \frac{dz}{2\pi i} J(z) \phi(w). \quad (2.212)$$

The integrand is a closed 1-form, so Q is independent of the contour C as long as it belongs to the given homology class. Hence we may deduce the Abelian Ward identities of the underlying $U(1)$ symmetry by standard contour manipulations as we did before for other conserved charges.

The OPE of $J(z)$ with itself must have the form

$$J(z) J(w) \sim \frac{k}{(z-w)^2} \quad (2.213)$$

³⁹ Replacing $J(z)$ with $\frac{1}{2}(J(z) + J(z)^\dagger)$ and $\frac{1}{2i}(J(z) - J(z)^\dagger)$, if necessary, we may (and do!) assume the current $J(z)$ to be self-adjoint.

for some positive constant k which we set to 1 as a choice of normalization of the operator $J(z)$. Indeed a pole of order 1 in the RHS of (2.213) is ruled out by the $z \leftrightarrow w$ symmetry of the LHS.⁴⁰

We bosonize the current $J(z)$ in terms of a scalar field $\phi(z)$, i.e. we write

$$J(z) = \partial\phi(z), \quad \text{with} \quad \phi(z) = \int^z dw J(w). \quad (2.214)$$

The OPE (2.213) is equivalent to

$$\phi(z)\phi(w) \sim -\log(z-w), \quad (2.215)$$

that is, $\phi(z)$ is a free chiral scalar with an action of the form (2.193). By a *chiral scalar* we mean a scalar field with only the left-moving part $\phi_L(z)$. A chiral scalar is a free field which satisfies the extra constraints $\bar{\partial}\phi(z) = 0$. While chiral scalars are free QFTs, their quantization is quite subtle because its constraints do not commute:

$$[\bar{\partial}\phi(z), \bar{\partial}\phi(z')] \neq 0. \quad (2.216)$$

In particular, being *chiral*, these scalars suffer from gravitational anomalies. Chiral scalars are important in string theory, and we shall return to them later in this chapter and in Chaps. 6, 7.

The Current Algebra of a Free Scalar

Returning to a non-chiral free scalar field $X(z, \bar{z})$, we see that the primary operators $\partial X(z)$ and $\bar{\partial} X(\bar{z})$ are a pair of conserved left- and right-moving chiral currents which generate an Abelian current algebra. For X a compact scalar the corresponding symmetry group G is compact, $G \equiv U(1)_L \times U(1)_R$, whereas G is non-compact for X non-compact. When X is a target-space coordinate of a string moving in flat spacetime, the associated Abelian symmetry is the group $\mathbb{R}$ of translations in the X -direction, and the charge is the corresponding component of momentum, p , which has a continuous spectrum (cf. Eq. (2.197)).

More precisely, the two charges associated with the two chiral currents are

$$P_L = \oint \frac{dz}{2\pi i} \partial X(z), \quad P_R = \oint \frac{d\bar{z}}{-2\pi i} \bar{\partial} X(\bar{z}) \quad (2.217)$$

which are integrally quantized in the compact case.

In the non-compact case we have the constraint

$$P_L - P_R = \frac{1}{2\pi i} \oint dX = 0 \quad (2.218)$$

⁴⁰ Here we are assuming that $J(z)$ is consistent with 2d *Spin and Statistics* and hence bosonic.

because the scalar field is univalued on the world-sheet. $P_L = P_R \equiv \sqrt{\alpha'/2}p$ then takes arbitrary real values (cf. (2.201)). Since spacetime translations correspond to the continuous global symmetry $X \rightarrow X + \text{const.}$ on the world-sheet, we conclude that *in string theory spacetime translations are never spontaneously broken by virtue of the 2d Coleman theorem [35] (see BOX 2.5).*

Exponentials of Free Scalars

The scalar theory has central charge $c = 1$, and hence infinitely many primaries by Cardy theorem. Besides the Abelian currents $\partial X, \bar{\partial} X$, the only other primary fields in this (non-compact) free CFT are the normal ordered exponentials

$$: \exp(ikX(w, \bar{w})), \quad (2.219)$$

where k is a real variable. They correspond to the non-trivial characters of the underlying non-compact Abelian group $\mathbb{R}$; from the target-space perspective they are plane waves of momentum k . The Hermitian conjugate to $: \exp(ikX(w, \bar{w})) :$ carries the dual character $: \exp(-ikX(w, \bar{w})) :$.

Exercise 2.8 Prove the following three OPEs

$$T(z) : e^{ikX(w, \bar{w})} := \left[\frac{\alpha' k^2}{4(z-w)^2} + \frac{\partial_w}{(z-w)} \right] : e^{ikX(w, \bar{w})} + \text{reg.} \quad (2.220)$$

$$\partial X(z) : e^{ipX(w, \bar{w})} := -\frac{\alpha'}{2} \frac{ip}{z-w} : e^{ipX(w, \bar{w})} : + \text{reg.} \quad (2.221)$$

$$: e^{ipX(z, \bar{z})} : : e^{iqX(w, \bar{w})} := |z-w|^{\alpha' p q} : e^{ipX(z, \bar{z}) + iqX(w, \bar{w})}. \quad (2.222)$$

Hint: Eq. (2.222) is known as the *Wick theorem*.

From the first OPE, we see that $: \exp(ikX(w, \bar{w})) :$ is a Virasoro primary of weight

$$h(k) = \frac{\alpha' k^2}{4}. \quad (2.223)$$

The spectrum of the radial Hamiltonian is continuous in this case. Indeed non-compact CFTs are not “good” CFT in the sense of the discussion around Eq. (2.38).

2.5 Free SCFTs and Their Bosonization

We extend the above discussion to *general* free conformal fields. We are particularly interested in the CFT describing the string/superstring chiral ghosts b, c, β, γ . The basic tool to dwell with such free theories is 2d *bosonization* [2, 38].

2.5.1 b, c and β, γ Systems

An important class of 2d CFTs is given by the **free** (linear) conformal theories. Focusing on the left-movers, we consider (Euclidean) actions of the form⁴¹

$$S = \frac{1}{2\pi} \int d^2z \, b \bar{\partial} c, \quad (2.224)$$

where the complex fields b, c are either *fermionic* (anticommuting) or *bosonic* (commuting); in the second case the fields are traditionally written using Greek letters β, γ . We refer to these CFTs as b, c systems and, respectively, β, γ systems. Unitarity is not an issue in this section, and most CFT below are *non-unitary*.

b, c, β, γ are conformal primaries as well as chiral operators, i.e. they have $\tilde{h} = 0$. Invariance of the action implies their conformal weights (equal to the conformal spins for chiral fields) to have the *complementary* values λ and $1 - \lambda$, that is,

$$b'(z') (dz')^\lambda = b(z) (dz)^\lambda, \quad c'(z') (dz')^{1-\lambda} = c(z) (dz)^{1-\lambda}. \quad (2.225)$$

The ghosts in the Polyakov quantization of the string/superstring form a b, c system with $\lambda = 2$ and a β, γ system with $\lambda = 3/2$.

In the geometric language b (resp. c) is a section of a holomorphic line bundle $\mathcal{L}$ which is a λ power (resp. a $(1 - \lambda)$ power) of the canonical bundle $K \rightarrow \Sigma$. In particular, when λ is half-integral, $\mathcal{L}$ is an odd power of a spin bundle; cf. **BOX 1.10**. The actions are (we focus on left-moving chiral fields)

$$S_{b,c} = \frac{1}{2\pi} \int d^2z \, b \bar{\partial} c, \quad S_{\beta,\gamma} = \frac{1}{2\pi} \int d^2z \, \beta \bar{\partial} \gamma, \quad (2.226)$$

from which we get the OPEs

$$b(z_1) c(z_2) \sim \frac{1}{z_1 - z_2} \quad c(z_1) b(z_2) \sim \frac{1}{z_1 - z_2} \quad (2.227)$$

$$\beta(z_1) \gamma(z_2) \sim -\frac{1}{z_1 - z_2} \quad \gamma(z_1) \beta(z_2) \sim \frac{1}{z_1 - z_2}. \quad (2.228)$$

The energy-momentum tensors are

$$T(z) = (\partial b)c - \lambda \partial(bc), \quad T(z) = (\partial \beta)\gamma - \lambda \partial(\beta\gamma). \quad (2.229)$$

Exercise 2.9 Use Eqs. (2.227)–(2.229) to check that $T(z)$ satisfies the CFT OPE (2.102) and compute the central charge c as a function of the weight λ .

⁴¹ We normalize the 2d volume form as $d^2z \equiv dz \wedge d\bar{z} = 2 dx \wedge dy$.

2.5.2 Anomalous $U(1)$ Current (“Ghost Number”)

A b, c system (resp. β, γ) has a chiral $U(1)$ symmetry. For the ghost systems the associated charges are called *ghost numbers* (left- and right-moving ghost numbers are separately conserved in the closed string). The (left-moving) ghost currents are conserved in flat space, but *on a general curved world-sheet they suffer a gravitational anomaly proportional to the world-sheet scalar curvature R* . The $U(1)$ currents are⁴²

$$j_c(z) = -bc, \quad j_\gamma(z) = -\beta\gamma, \quad (2.230)$$

with OPEs

$$j_c(z) c(w) \sim \frac{c(w)}{z-w} \quad j_c(z) b(w) \sim -\frac{b(w)}{z-w} \quad (2.231)$$

$$j_\gamma(z) \gamma(w) \sim \frac{\gamma(w)}{z-w} \quad j_\gamma(z) \beta(w) \sim -\frac{\beta(w)}{z-w}, \quad (2.232)$$

which just say that c and γ have charge $+1$ while b and β have charge -1 .

The two-current OPE are

$$j_c(z) j_c(w) \sim \frac{1}{(z-w)^2}, \quad j_\gamma(z) j_\gamma(w) \sim \frac{-1}{(z-w)^2}. \quad (2.233)$$

The anomaly of the current is manifest in its OPE with the energy–momentum tensor

$$T(z) j(w) \sim \frac{Q}{(z-w)^3} + \frac{j(w)}{(z-w)^2} + \frac{\partial j(w)}{z-w} \quad (2.234)$$

where $j(w)$ stands for either $j_c(w)$ or $j_\gamma(w)$ and

$$Q = \begin{cases} (1-2\lambda) & b, c \text{ system} \\ -(1-2\lambda) & \beta, \gamma \text{ system.} \end{cases} \quad (2.235)$$

For a free system *globally defined* on all Σ 's, one has $2\lambda \in \mathbb{Z}$, so $Q \in \mathbb{Z}$.

Exercise 2.10 Check Eqs. (2.234), (2.235).

Exercise 2.11 Prove the following formula for the Virasoro central charge c :

$$c = \epsilon(1 - 3Q^2), \quad \text{where } \epsilon = \begin{cases} +1 & \text{for } b, c \\ -1 & \text{for } \beta, \gamma. \end{cases} \quad (2.236)$$

⁴² **Beware:** To simplify the notation, in this section we mostly omit the symbol of conformal normal order: e.g. bc actually stands for $:bc:$.

Deferred Proof of Claim 1.1 from Chap. 1

From **BOX 2.3** the coefficient of the Weyl anomaly is $-c/12$ (for left-movers). For free chiral fields of conformal spin λ , the central charge c is given by the formula (2.236). This proves the **clm**.

From (2.234) we see that the current $j(z)$ is *not* a conformal primary.⁴³ In terms of modes Eq. (2.234) reads

$$[L_m, j_n] = -n j_{m+n} + \frac{1}{2} Q m(m+1) \delta_{m+n,0}. \quad (2.237)$$

For the zero-mode this equation implies

$$j_0^\dagger = -[L_{-1}, j_1]^\dagger = -[L_1, j_{-1}] = -j_0 - Q, \quad (2.238)$$

so Q may be interpreted as a background charge on the sphere, as we shall check momentarily. We stress that $j(z)$ is not even quasi-primary:⁴⁴

$$L_1|j\rangle = L_1 j_{-1}|0\rangle = [L_1, j_{-1}]|0\rangle = Q|0\rangle. \quad (2.239)$$

2.5.3 Fermi/Bose Sea States

The energy of the free linear systems (2.226) is *unbounded below*, so the usual notion of ground state is useless. In the Fermi case, the standard Dirac prescription is to use the *Fermi sea states* in which all energy levels below a certain energy are filled. For the β, γ system we need the corresponding notion, the *Bose sea states*, which however should involve a much subtler construction since in Bose statistics we cannot simply “fill” the levels: indeed, no matter how large the occupation number of a level, there remains plenty of room to insert more “stuff” in that same level.

That the bosonic case should be rather subtle can be understood physically as follows: as explained in Sect. 1.6.1, the chiral free systems (2.226) may be regarded as a Weyl spinor coupled to a certain background $U(1)$ gauge field A . Therefore, if we insist that the fields are bosonic, we are violating *by hand* the 2d *Spin and Statistics* theorem. The original proof by Pauli of the theorem was based on the observation that quantizing a spinor field using Bose statistics introduces “pathologies” incompatible with a sound QFT (see p. 722 of [39]). Here we dare construct a consistent quantization of the bosonic β, γ system in open rebellion against the most sacred principles. The “price” we pay for our defiance is that its quantization will look

⁴³ That is, $j(z)$ is not a genuine 1-form. Hence we cannot prove charge conservation on a general world-sheet by usual contour manipulations: this fact signals the existence of a mixed gravitational- $U(1)$ anomaly. Thus Q measures the failure of current conservation in curved world-sheets.

⁴⁴ Quasi-primary chiral currents lead to non-anomalous Ward identities by the “good atlas” argument after Eq. (2.61). To have an anomalous Ward identity the current j must be *non*-quasi-primary. This also entails that $L_1|j\rangle$ is a measure of the anomaly.

a bit “strange”. It is a very welcomed price, since it will have wonderful physical implications. Despite being a little unusual, our procedure below is well-defined.

Mode Expansions

For definiteness we write the following formulae assuming $\lambda \in \mathbb{Z}$ for the b, c system, and $\lambda \in \mathbb{Z} + \frac{1}{2}$ for β, γ one: these are the most relevant cases. The extension to other values of λ for either statistics is straightforward and left to the reader as an easy exercise. Then for the β, γ system we have two possible choices of spin-structure on the cylinder $S^1 \times \mathbb{R}$: along the circle S^1 the spinorial fields β, γ may be

- *periodic*: to be called the *Ramond sector* (R),
- *anti-periodic*: to be called the *Neveu–Schwarz sector* (NS),

that is, we have the periodic b.c.

$$\begin{aligned} \beta(w + 2\pi i) &= -e^{2\pi i\alpha} \beta(w) \\ \gamma(w + 2\pi i) &= -e^{2\pi i\alpha} \gamma(w) \end{aligned} \quad \text{where } \alpha = \begin{cases} \alpha = \frac{1}{2} & \text{R sector} \\ \alpha = 0 & \text{NS sector.} \end{cases} \quad (2.240)$$

On the plane (in radial quantization) a field with $\lambda \in \frac{1}{2} + \mathbb{Z}$ gets multiplied by the phase $e^{2\pi i\alpha}$ when going around the origin; cf. Eq. (2.86).

To define the *sea states*, we consider the mode expansions

$$b(z) = \sum_{n \in \mathbb{Z} - \lambda} \frac{b_n}{z^{n+\lambda}} \quad c(z) = \sum_{n \in \mathbb{Z} + \lambda} \frac{c_n}{z^{n+1-\lambda}} \quad (2.241)$$

$$\beta(z) = \sum_{n \in \mathbb{Z} - \lambda + \alpha} \frac{\beta_n}{z^{n+\lambda}} \quad \gamma(z) = \sum_{n \in \mathbb{Z} + \lambda + \alpha} \frac{c_n}{z^{n+1-\lambda}}, \quad (2.242)$$

where the modes satisfy⁴⁵

$$c_n^\dagger = c_{-n} \quad b_n^\dagger = b_{-n} \quad (2.243)$$

$$\gamma_n^\dagger = \gamma_{-n} \quad \beta_n^\dagger = -\beta_{-n} \quad (2.244)$$

$$c_m b_n + b_n c_m = \delta_{m+n,0}, \quad \gamma_m \beta_n - \beta_n \gamma_m = \delta_{m+n,0}. \quad (2.245)$$

A *Fermi/Bose sea* is a state $|q\rangle$ which splits the normal modes

⁴⁵ From the general CFT formulae, if L_m are the modes of the energy–momentum tensor, we have

$$[L_m, b_n] = -(- (1 - \lambda)m - n)b_{m+n}, \quad [L_m, c_n] = -(- \lambda m - n)c_{m+n}$$

and the same formulae with $b_n \rightarrow \beta_n, c_n \rightarrow \gamma_n$.

$$\begin{array}{ll} \text{Fermi sea} & \begin{cases} b_n|q\rangle = 0 & n > q - \lambda \\ c_n|q\rangle = 0 & n \geq -q + \lambda \end{cases} \quad (2.246) \\ \text{Bose sea} & \begin{cases} \beta_n|q\rangle = 0 & n > -q - \lambda \\ \gamma_n|q\rangle = 0 & n \geq q + \lambda. \end{cases} \quad (2.247) \end{array}$$

The number $q \in \alpha + \mathbb{Z}$ which labels the sea state is called the *sea level*. Inserting

$$1 = c_m b_{-m} + b_{-m} c_m \quad (2.248)$$

in $\langle q'|q\rangle$ we see that the only non-zero inner product is⁴⁶

$$\langle -q - Q|q\rangle = 1, \quad (2.249)$$

a result that we shall re-interpret momentarily (in more than one way). Thus

$$\begin{aligned} \langle c(z)b(w)\rangle_q &\equiv \langle -q - Q|c(z)b(w)|q\rangle = \\ &= \sum_{m,n} \frac{1}{z^{m+1-\lambda}} \frac{1}{w^{n+\lambda}} \langle -q - Q|c_m b_n|q\rangle = \\ &= \sum_{\substack{m \geq -q+\lambda \\ n \leq q-\lambda}} \frac{1}{z^{m+1-\lambda}} \frac{1}{w^{n+\lambda}} \delta_{m+n,0} = \left(\frac{z}{w}\right)^q \frac{1}{z-w}, \end{aligned} \quad (2.250)$$

and similarly⁴⁷

$$\langle \gamma(z)\beta(w)\rangle_q = \left(\frac{w}{z}\right)^q \frac{1}{z-w}. \quad (2.251)$$

We stress that only the sea state $|q=0\rangle$ is translational invariant. Then

$$\langle j(z)\rangle_q = \lim_{w \rightarrow z} \left\langle c(z)b(w) - \frac{1}{z-w} \right\rangle_q = \frac{1}{z} \lim_{w \rightarrow z} \frac{(z/w)^q - 1}{z/w - 1} = \frac{q}{z}, \quad (2.252)$$

so that

$$j_0|q\rangle = q|q\rangle, \quad (2.253)$$

and Eq. (2.249) follows from (2.238). Equations (2.252), (2.253) hold also for β, γ .

We note that (2.249) implies that Q is the background charge on the sphere S^2 . Indeed, we already saw that the “ghost number” current $j(z)$ has an anomaly proportional to the curvature of the world-sheet: Q measures the amount of non-conservation of the charge for the free CFT defined on the sphere. For more details on this issue, see the next subsection.

⁴⁶ The overall normalization is, of course, conventional.

⁴⁷ Note that Eq. (2.247) is obtained from (2.246) by the formal substitution $q \rightarrow -q$.

Using Eq. (2.229), for a b, c system we get⁴⁸

$$\begin{aligned} \langle T(z) \rangle_q &= \lim_{w \rightarrow z} \left(\left((\lambda - 1) c(z) \partial_w b(w) + \lambda \partial_z c(z) b(w) - \text{sing.} \right) \right)_q = \\ &= \lim_{w \rightarrow z} \left\{ (\lambda - 1) \partial_w + \lambda \partial_z \right\} \left\{ \left[\left(\frac{z}{w} \right)^q - 1 \right] \frac{1}{z - w} \right\} = \frac{q(Q + q)}{2z^2}. \end{aligned} \quad (2.254)$$

Going through the sign flips which relate the expressions for β, γ to the corresponding ones for b, c , we see that for a β, γ system

$$\langle T(z) \rangle_q = -\frac{q(Q + q)}{2z^2}. \quad (2.255)$$

Thus

$$L_0 |q\rangle = \epsilon \frac{1}{2} q(q + Q) |q\rangle, \quad \epsilon = \begin{cases} +1 & \text{Fermi} \\ -1 & \text{Bose.} \end{cases} \quad (2.256)$$

Exercise 2.12 Check that $L_n |q\rangle = 0$ for $n > 0$.

We conclude that $|q\rangle$ is a primary state of weight $h = \epsilon q(q + Q)/2$.

2.5.4 The $U(1)$ Stress Tensor and Its Bosonization

We define the $U(1)$ stress tensor as

$$T^j(z) \stackrel{\text{def}}{=} \frac{1}{2} \epsilon (j(z)^2 - Q j(z)) \quad Q \equiv \epsilon(1 - 2\lambda), \quad (2.257)$$

where, to simplify the notation, we omit writing the symbol of normal product (defined, as always, by subtracting the OPE singularities). The linear term in Eq. (2.257) is designed so that $T(z)$ and $T^j(z)$ have the same commutation relations with $j(w)$, i.e. the same singular part of the OPE:

$$T^j(z) j(w) = T(z) j(w) + \text{regular as } z \rightarrow w. \quad (2.258)$$

On the other hand, a simple computation yields

$$T^j(z) T^j(w) \sim \frac{1}{2} \frac{c^j}{(z - w)^4} + \frac{2 T^j(w)}{(z - w)^2} + \frac{\partial_w T^j(w)}{z - w}, \quad (2.259)$$

⁴⁸ Indeed, $\frac{e^{q \log(z/w)} - 1}{z - w} = \frac{q(\log z - \log w) + \frac{q^2}{2}(\log z - \log w)^2}{z - w} + O((z - w)^2) = \frac{q}{w} + \frac{1}{2} q(q - 1) \frac{z - w}{w^2} + O((z - w)^2).$

where

$$c^j = 1 - 3\epsilon Q^2 = \begin{cases} c & \text{Fermi} \\ c + 2 & \text{Bose.} \end{cases} \quad (2.260)$$

Exercise 2.13 Check Eqs. (2.258) and (2.259), (2.260).

From Eq. (2.259) we see that $T^j(z)$ is a conformal energy–momentum tensor on its own right, whose Virasoro central charge is c^j . Consider now the local operator

$$T'(z) \stackrel{\text{def}}{=} T(z) - T^j(z). \quad (2.261)$$

By the very construction of $T^j(z)$, the OPE of $T'(z)$ with $j(z)$ is non-singular

$$T'(z) j(w) \equiv \left(T(z) - T^j(z) \right) j(w) \sim 0. \quad (2.262)$$

Since $T^j(w)$ is a polynomial in $j(w)$ and its derivatives, (2.262) also entails that

$$T'(z) T^j(w) \equiv \left(T(z) - T^j(z) \right) T^j(w) \sim 0. \quad (2.263)$$

Hence

$$T'(z) T'(w) \sim \left(T(z) - T^j(z) \right) T(w) \sim \frac{c - c^j}{(z - w)^4} + \frac{2 T'(w)}{(z - w)^2} + \frac{\partial T'(w)}{z - w}, \quad (2.264)$$

so $T'(z)$ is also a valid CFT energy–momentum tensor with central charge

$$c' \equiv c - c^j. \quad (2.265)$$

An indecomposable CFT has a *unique* energy–momentum tensor. We conclude that our free CFT decomposes into two *non-interacting* CFT: the **anomalous $U(1)$ current algebra CFT** with energy–momentum tensor $T^j(z)$ and central charge c^j , and a “residual” CFT with energy–momentum tensor $T'(z)$ and central charge c' .

From Eq. (2.260) we see that in the Fermi case $c' = 0$. In view of Eq. (2.154), this means that $T'(z) = 0$ as an operator.⁴⁹ Therefore, *in the Fermi case* the $U(1)$ current algebra CFT is *equivalent* to the free b, c system. In other words, all operators of the b, c system may be written in terms of the chiral scalar field which bosonizes the anomalous current (see below). This equivalence of CFTs is a generalization of the well-known *bosonization* procedure for 2d fermions [40, 41]⁵⁰ which corresponds to the anomaly-free case $Q = 0$, that is, to complex fermions of spin $\frac{1}{2}$.

On the contrary, from Eq. (2.260) we see that in the *Bose case* the “residual” CFT has central charge $c' = -2$: therefore it must be non-trivial and **non-unitary**.

⁴⁹ Since the theory is non-unitary one should be more pedantic and say that the corresponding state $|T'\rangle$ has zero norm, and may be killed by the prescription of getting rid of null states.

⁵⁰ See **Appendix 1** to this chapter for Witten’s *non-Abelian* 2d bosonization procedure.

The “residual” $c = -2$ conformal system will be described in Sect. 2.5.6. Before going to that, we complete the deflist of the bosonization procedure for the anomalous $U(1)$ current algebra and for the related fermionic b, c theory.

Bosonization of the $U(1)$ Current

As in Sect. 2.4, we write the $U(1)$ charge in terms of a *chiral* scalar $\phi(z)$

$$j(z) = \epsilon \partial \phi(z), \quad \phi(z) = \epsilon \int^z dw j(w). \quad (2.266)$$

The OPEs (2.233) yield⁵¹

$$j(z) \phi(w) \sim \frac{1}{z-w}, \quad \phi(z) \phi(w) \sim \epsilon \log(z-w), \quad (2.267)$$

$$j(z) e^{q\phi(w)} \sim \frac{q}{z-w} e^{q\phi(w)}, \quad e^{q\phi(z)} e^{q'\phi(w)} = (z-w)^{\epsilon qq'} e^{q\phi(z)+q'\phi(w)}. \quad (2.268)$$

Exercise 2.14 Check the following OPE

$$T^j(z) e^{q\phi(w)} \sim \frac{\frac{1}{2}\epsilon q(q+Q)}{(z-w)^2} e^{q\phi(w)} + \frac{1}{z-w} \partial_w e^{q\phi(w)}. \quad (2.269)$$

The above OPEs say that $e^{q\phi(z)}$ is a *primary* conformal operator of the anomalous $U(1)$ CFT with $U(1)$ charge q and Virasoro weight

$$h = \frac{1}{2}\epsilon q(q+Q). \quad (2.270)$$

Comparing with the charges and weights computed in Sect. 2.5.3 for the sea state $|q\rangle$, we conclude that the soliton⁵² operator $e^{q\phi}$ shifts the Fermi/Bose sea level by q units

$$e^{q\phi(0)}|0\rangle = |q\rangle. \quad (2.271)$$

Equation (2.271) is the state-operator correspondence

$$e^{q\phi(z)} \longleftrightarrow |q\rangle. \quad (2.272)$$

In particular we see that

$$\langle 0|e^{-Q\phi(0)}|0\rangle = 1, \quad (2.273)$$

i.e. to get a non-zero result we need to insert an operator of charge $-Q$ to adsorb the background charge Q on the sphere (cf. Eq. (2.249)).

⁵¹ Recall that in this section the normal order symbols are left implicit: $e^{q\phi(w)}$ stands for $: e^{q\phi(w)} :$.

⁵² $e^{\pm\phi}$ is the Mandelstam soliton (or disorder) operator in old-fashioned 2d bosonization [41].

BOX 2.6 - Ghost current anomaly versus Riemann–Roch theorem

Equation (2.249) (or (2.273)) says that the $U(1)$ chiral anomaly integrated over the sphere is

$$\begin{aligned} 2\lambda - 1 = -Q &= \#(c \text{ zero-modes on } S^2) - \#(b \text{ zero-modes on } S^2) = \\ &= \dim H^0(S^2, K^{1-\lambda}) - \dim H^0(S^2, K^\lambda) = 1 - g(S^2) + \deg K^{1-\lambda} \end{aligned}$$

More generally $\bar{\partial} j(z)$ should be proportional to the scalar curvature R . Integrating both sides and using the Gauss–Bonnet formula (1.10) we get

$$\bar{\partial} j = -\frac{Q}{8} \sqrt{g} R.$$

Then the integral over a genus g surface yields

$$\dim H^0(\Sigma_g, K^{1-\lambda}) - \dim H^0(\Sigma_g, K^\lambda) = (2\lambda - 1)(g - 1) = 1 - g + \deg K^{1-\lambda}$$

which is the Riemann–Roch theorem.

Bosonization of b, c Fermi Fields

For Fermi statistics we have the bosonization formulae

$$b(z) = e^{-\phi(z)}, \quad c(z) = e^{\phi(z)}, \quad (2.274)$$

as we check by comparing charges and weights of the two sides as well as their OPEs. In fact, as we have explained after Eq. (2.265), the $\phi(z)$ chiral field with a background charge Q and the b, c system are equivalent.

For β, γ the story must be subtler, since these fields are bosons, while the soliton operators $e^{\mp\phi(z)}$ are fermions.

2.5.5 Riemann–Roch and Bosonization: The Linear Dilaton CFT

The anomalies, which are quantum effects in the fermionic formulation, are canonical equations of motion in the bosonized version of the theory, i.e. a tree-level effect.

We write the current $j = \partial\phi$ for a scalar field ϕ . The anomalous conservation of the $U(1)$ charge (BOX 2.6) is the canonical e.o.m. of the scalar ϕ with action

$$\frac{1}{4\pi} \int d^2z \left(-\partial\phi\bar{\partial}\phi + \frac{Q}{\pi} R\phi \right) \quad (2.275)$$

which has the form of a free scalar in a linear dilaton background; cf. Eq. (1.163) and the discussion around it.

However here there is a major subtlety. The b, c system is a *chiral* CFT, meaning that it is purely left-moving. So its bosonic equivalent is (2.275) with its right-moving d.o.f. suppressed, that is, with the additional *constraint*

$$\bar{\partial}\phi = 0 \quad (2.276)$$

(cf. discussion around Eq. (2.216)). The standard (i.e. non-chiral) scalar linear dilaton CFT with action (2.275) is then equivalent to the (non-chiral) left-right symmetric combination of free systems

$$\frac{1}{2\pi} \int \left(b\bar{\partial}c + \tilde{b}\partial\tilde{c} \right) \quad (2.277)$$

with the same value of the spin λ on the left and on the right.

Exercise 2.15 Check that the central charge of the linear dilaton system (which is still a free CFT) is $c = 1 - 3Q^2$, that is, equal to the central charge of the equivalent b, c system, Eq. (2.236).

The above bosonized form of the action, Eq. (2.275), is also very convenient to compute the partition function of the model (2.277) which, in the bosonic form, is a straightforward Gaussian integral without subtleties from quantum anomalies.

When the world-sheet Σ has higher genus, the bosonization requires some extra care with global aspects, which are also well understood in the literature; see [42].

2.5.6 Bosonization of β, γ : The $c = -2$ System

From the arguments in Sect. 2.5.4, we know that the “bosonization” of a Bose β, γ system involves, besides the chiral scalar ϕ associated with its anomalous $U(1)$ current algebra, a $c = -2$ CFT which must be a left-moving chiral *free* theory (since the full theory was free to start with) whose fields should be *anticommuting* to produce the right statistics for the bosonic fields β, γ .

The η, ξ and η, ρ systems

An anticommuting, chiral, free CFT is a Fermi b, c system. We impose that its central charge is -2

$$-2 = c \equiv (1 - 3Q^2) \Rightarrow Q = \pm 1 \Rightarrow \lambda = 1, 0. \quad (2.278)$$

The $\lambda = 1$ b, c system (equivalent to the $\lambda = 0$ one) is just a free complex (chiral) scalar with the *wrong* statistics. Its central charge is then *minus* the central charge of *two* copies of the real free scalar, so $c = -2$.

This peculiar system deserves special names, and we write $\eta(z)$ and $\xi(z)$ for its Fermi fields replacing the generic notation $b(z), c(z)$. As before, we have

$$\eta(z) \xi(w) \sim \frac{1}{z-w} \sim \xi(z) \eta(w). \quad (2.279)$$

Thus a β, γ system is equivalent to a free chiral scalar ϕ with background charge Q (as described in Sect. 2.5.5) together with the free η, ξ Fermi system.

Comparing $U(1)$ charge, weight, and OPEs, we get the “bosonization” formulae⁵³

$$\beta(z) = e^{-\phi(z)} \partial \xi(z), \quad \gamma(z) = e^{\phi(z)} \eta(z). \quad (2.280)$$

For instance, let us check the weights

$$h_\beta = \frac{1}{2} \left(- (1 - 2\lambda) - 1 \right) + 1 \equiv \lambda, \quad (2.281)$$

$$h_\gamma = -\frac{1}{2} \left(- (1 - 2\lambda) + 1 \right) + 1 \equiv 1 - \lambda. \quad (2.282)$$

The $\eta(z), \xi(z)$ fermionic system has its own $U(1)$ chiral current $-\eta\xi(z)$ (as before), and it may be bosonized in terms of a second chiral scalar $\chi(z)$

$$\xi(z) \eta(z) = \partial \chi(z), \quad \chi(z) \chi(w) \sim \log(z-w), \quad (2.283)$$

$$\eta = e^{-\chi}, \quad \xi = e^{\chi}. \quad (2.284)$$

A crucial remark is that, since $\xi(z)$ is a 0-form, it has precisely *one* zero-mode on every connected world-sheet Σ , namely the *constant mode*. The zero-mode algebra

$$\eta_0^2 = \xi_0^2 = 0, \quad \{\eta_0, \xi_0\} = 1 \quad (2.285)$$

forces the ground state of the ξ, η system to be *two-fold* degenerate. The two ground states are the $SL(2, \mathbb{C})$ -invariant state $|0\rangle$ and its Hermitian conjugate $|-Q\rangle \equiv |1\rangle$ which carry different charges because of the anomaly; see Eq. (2.238).

Since $\xi = e^\chi$, from Eq. (2.249) (with $q = 1$)

$$\langle 0 | \xi_0 | 0 \rangle = \langle 0 | \xi(z) | 0 \rangle = \langle 0 | e^{\chi(0)} | 0 \rangle = \langle 0 | 1 \rangle = 1. \quad (2.286)$$

In the path integral language this equation reflects the need of the insertion of a Fermi field $\xi(z)$ to “soak up” its zero-mode to get a non-zero result on the sphere.⁵⁴

⁵³ There are other possible choices for the bosonization formulae; see **BOX 2.7**. The ones in the text are the most common ones in the superstring literature [2].

⁵⁴ This “soaking up” procedure for fermionic zero-modes should be familiar to the reader from QCD in the instanton background [43, 44].

BOX 2.7 - An alternative bosonization of the β, γ system

Equation (2.280) are the Friedan–Matinec–Shenker (FMS) bosonization formulae for the β, γ system [2]; this is the bosonization scheme which is standard in the superstring literature. There is an alternative bosonization [45] in the form

$$\gamma = -e^\phi \partial \xi, \quad \beta = e^{-\phi} \eta.$$

The two versions have equivalent “small” algebras but are inequivalent at the “large” algebra level. The second version has two advantages: *i*) it allows a manifestly supersymmetric bosonization of the supersymmetric b, c, β, γ system, and *ii*) in its application to the superstring the BRST cohomology in the “large” algebra is simply related to the physical states. However it is less convenient than the FMS scheme for actual computations. Therefore in the rest of this textbook we shall use only the FMS version.

“Small” and “Big” Operator Algebras

However the original fields β, γ contain only $\rho \equiv \partial \xi$ and not ξ itself. Thus the relevant $c = -2$ system is really given by the two $(1, 0)$ currents η and ρ , rather than by η and ξ . In other words: **the zero-mode ξ_0 of $\xi(z)$ is not part of the original operator algebra.**⁵⁵

In the $\eta(z), \rho(z)$ chiral algebra the operator η_0 is then *central*, so a complex number in any irreducible representation; then it is consistent to fix $\eta_0 = 1$ as a “choice of normalization”. The reduced η, ρ system has a *unique* ground state $|0\rangle_{\eta, \rho}$, and

$$\langle 0 | \cdots | 0 \rangle_{\eta, \rho} = \langle 0 | \xi_0 \cdots | 0 \rangle_{\eta, \xi}. \quad (2.287)$$

We shall refer to the system $e^{\pm\phi}, \eta, \rho$ as the “*small*” system and to the one with $e^{\pm\phi}, \eta, \xi$ as the “*large*” one. The “small” system has a unique vacuum, while the “large” one has a doubly degenerate vacuum.

2.5.7 The Picture Charge

In the bosonization of β, γ systems there is a crucial issue which we must stress. While in the original field content

$$\beta(z), \quad \gamma(z), \quad (2.288)$$

there was a single $U(1)$ charge, associated with the current $-\beta(z)\gamma(z)$ (which we called the “ghost number”), in the bosonized setup with fields

⁵⁵ Stated differently: the “small” operator algebra is the algebra of CFT operators (primaries and descendants) of a wrong statistics complex scalar, which does not contain the zero-mode.

$$\phi(z), \quad \xi(z), \quad \eta(z), \quad (2.289)$$

we have an additional quantum number, i.e. the $U(1)$ charge of the ξ, η system. The original bosonic fields β, γ , Eq. (2.280), are invariant with respect to the difference of the ξ, η charge and the $\partial\phi$ charge⁵⁶

$$N_p = \oint \frac{dz}{2\pi i} (\xi\eta - \partial\phi) = \oint \frac{dz}{2\pi i} (\partial\chi - \partial\phi), \quad (2.290)$$

so this quantum number is not visible in the original formulation. We write n_p for the value of the charge N_p . This new quantum number is called the *picture charge*.

Discussion

Jumping a bit ahead, we outline the relevance of the picture charge for superstring theory. Most constructions/computations in the superstring require a bosonization of the β, γ ghost system. In the bosonized formulation, the states are then classified by an additional quantum number, the picture charge n_p . The Hilbert space may be decomposed into subspaces of definite picture charge

$$\mathbf{H} = \bigoplus_{n_p} \mathbf{H}_{n_p}. \quad (2.291)$$

Naively one would expect that this new quantum number is trivial (i.e. that only the zero picture Hilbert space $\mathbf{H}_0$ matters for physical processes), but this cannot be possibly true since the picture current inherits a 2d gravitational anomaly from the β, γ and η, ξ currents (see Sect. 2.5.2) and hence on a curved world-sheet Σ there is a background picture charge which depends on its topology *via* the Riemann–Roch theorem (BOX 2.6). This implies that *the picture charge is **not** conserved*⁵⁷ in superstring perturbation theory, and we may not fix it to be zero once and for all!

On the other hand, physical observables *should not* depend on the picture charge assignments of the states, since this quantum number is, in a quite strong sense, artificial and “unphysical”. *Is there a way out of this conundrum?*

Yes, there is one, albeit quite subtle. First of all, notice that the Hilbert space $\mathbf{H}$ of the “matter + ghost” world-sheet CFT cannot be the space of physical states of the superstring since its Hermitian product is not positive-definite (because the world-sheet CFT is non-unitary due to the presence of ghosts and longitudinal modes). The physical Hilbert space $\mathbf{H}_{\text{phy}}$ is then a sub-quotient of $\mathbf{H}$ defined by the BRST quantization procedure (see Chap. 3). The BRST charge is canonically constructed using the Faddeev–Popov ghosts β, γ , and so has zero picture charge. We can then consider the Hilbert space of *physical states* with a given picture-charge n_p

$$(\mathbf{H}_{\text{phy}})_{n_p} \subset \mathbf{H}_{\text{phy}}. \quad (2.292)$$

⁵⁶ Recall that the $\partial\phi$ charge is *minus* the β, γ ghost current.

⁵⁷ More precisely, n_p is conserved only mod 1. So we can limit ourselves to the subsector of the formal Hilbert space with $n_p = \mu + \mathbb{Z}$ for a fixed μ .

Suppose the initial state of our physical process belongs to some $(\mathbf{H}_{\text{phy}})_{n_p}$. Since n_p is conserved only mod 1, the final state will be in $(\mathbf{H}_{\text{phy}})_{n'_p}$ with $n'_p \neq n_p$. The only way that the process may be independent of n_p is that the Hilbert spaces of *physical* states $(\mathbf{H}_{\text{phy}})_{n_p}$ are *isomorphic* for all values of n_p or, more precisely, that they depend (up to isomorphism) only on the class of n_p mod 1. If this “unlikely miracle” turns out to happen (as we shall show), we can perform our computations of physical observables using representatives of states in any one of the $(\mathbf{H}_{\text{phy}})_{n_p}$ ’s and get the same answer. States in $(\mathbf{H}_{\text{phy}})_{n_p}$ are said to have “ n_p picture”, while replacing the physical states in one space $(\mathbf{H}_{\text{phy}})_{n_p}$ with the corresponding ones in the isomorphic space $(\mathbf{H}_{\text{phy}})_{n'_p}$ is called *picture changing*. The anomaly then implies that if we choose the **|in>** states to be represented by vectors in some picture, the **<out|** states will be vectors in some *other* picture; however we can use the isomorphism to rewrite them in the original picture. The physical amplitude so defined should be independent of all picture choices if the theory has to be consistent.

Pictures versus “Soaking-Up” Zero-Modes

Changing the sea level q in the Fermi b, c system leads to an *equivalent* Fock space representation of the b, c chiral algebra, that is, the Fock space constructed by acting with the modes b_m, c_n on the Fermi sea $|q\rangle$ is independent of q . This is what we really mean when we say that we pass from $|q\rangle$ to $|q'\rangle$ by filling/emptying a suitable (finite) set of energy levels.

As mentioned after Eq. (2.286), this equivalence of Fock spaces has an important consequence for the path integral formulation, namely the prescription to “soak up” the Fermi zero-modes. Recall that, in view of the Riemann–Roch theorem, equation (2.273) just says that on the sphere $\Sigma \equiv S^2$ there are $-Q \equiv 2\lambda - 1$ (net) zero-modes of the Fermi field $c(z)$. In the operator formalism, to get a non-zero amplitude, we have to convert the Fermi sea $|0\rangle$ to the Fermi sea $|-Q\rangle$ by acting on it with $-Q$ modes of the c -field, that is,⁵⁸

$$|-Q\rangle = c_{\lambda-1} c_{\lambda-2} c_{\lambda-3} \cdots c_{-\lambda+1} |0\rangle = \kappa c(\partial c)(\partial^2 c) \cdots (\partial^{-1-Q} c) |0\rangle. \quad (2.293)$$

Correspondingly, in the path integral, to get a non-zero answer, we have to insert a number $-Q$ of c fields to *soak up* the zero-modes of the Fermi field

$$\begin{aligned} \langle -Q | \mathcal{O}_1(z_1) \cdots \mathcal{O}_s(z_s) | 0 \rangle & \stackrel{\text{operator}}{\underset{\text{formulation}}{=}} \\ & = \kappa \left\langle \mathcal{O}_1(z_1) \cdots \mathcal{O}_s(z_s) c(\partial c)(\partial^2 c) \cdots (\partial^{-1-Q} c) \right\rangle_{\text{path}} \neq 0. \end{aligned} \quad (2.294)$$

This simple story does not apply to the bosonic β, γ chiral algebra. Now one cannot change the value of the level q by acting on the Bose sea state $|q\rangle$ with *finitely many* oscillator modes β_m, γ_m : it is impossible to “fill” bosonic levels! Bose seas are *coherent* states not Fock states, and the Fock spaces built over different Bose seas

⁵⁸ In (2.293) κ is the obvious normalization factor.

$|q\rangle, |q'\rangle$ are *inequivalent*, that is, they yield unitarily *inequivalent* realizations of the free canonical commutator relations (2.245).

The situation is more transparent in the path integral language: bosonic zero-modes make the path integral *divergent* rather than vanishing, unless they correspond to collective coordinates taking value in *compact* manifolds. This yields a heuristic understanding of the zero-mode divergence as arising from an overall infinite volume in field space. Then there cannot be any operator polynomial in (derivatives of) β, γ whose insertion makes the bosonic integral finite and non-zero.

“Soaking-up” bosonic zero-modes is a subtle story. To shed light, we may use a *formal* analogy with the fermionic case. Insertion of c is actually insertion of $\delta(c)$ according to the rules of Grassmanian integration. Then we expect that to cure the β, γ zero-modes we have to insert in the path integral operators of the form

$$\delta(\gamma(z)) \equiv \int \frac{dk}{2\pi} e^{ik\gamma(z)} \quad \text{or} \quad \delta(\beta(z)) \equiv \int \frac{dk}{2\pi} e^{ik\beta(z)}. \quad (2.295)$$

Indeed, in the presence of zero-modes of $\gamma(z)$, or respectively of $\beta(z)$, these insertions make the path integral finite since they fix the value of the non-compact collective coordinates. From the integral representation in the RHS of (2.295), one finds

$$\beta(z) \delta(\beta(w)) \sim (z - w) \partial\beta(w) \delta(\beta(w)) \quad (2.296)$$

$$\gamma(z) \delta(\beta(w)) \sim \frac{1}{(z - w)} \delta'(\beta(w)), \quad (2.297)$$

and similar formulae for $\delta(\gamma(z))$. Note that $\delta(\beta(z))$ (resp. $\delta(\gamma(z))$) carries the *opposite* quantum numbers with respect to $\beta(z)$ (resp. $\gamma(z)$).

Exercise 2.16 Prove the above OPEs.

Correspondingly, in the operator formalism, the action of the modes of $\delta(\beta(z))$, $\delta(\gamma(z))$ on the Bose sea states is (see also [46])

$$\delta(\beta_{-q-\lambda}) |q\rangle = |q + 1\rangle \quad \delta(\gamma_{q+\lambda-1}) |q\rangle = |q - 1\rangle. \quad (2.298)$$

Therefore, we have the state-operator correspondence

$$\delta(\beta(z)) \leftrightarrow |q = +1\rangle, \quad \delta(\gamma(z)) \leftrightarrow |q = -1\rangle. \quad (2.299)$$

But we already know from Eq. (2.271) that the Bose sea state to $|q\rangle$ corresponds to the operator $e^{q\phi(z)}$. Hence we get the bosonization formulae

$$\delta(\beta(z)) \simeq e^{\phi(z)}, \quad \delta(\gamma(z)) \simeq e^{-\phi(z)}. \quad (2.300)$$

Note, however, that two sides differ in picture charge. That is, while the bosonization rule (2.300) is correct, implicitly it involves choices. This was to be expected: since

we have infinitely many inequivalent realizations of the β, γ operator algebra, we have infinitely many inequivalent ways to realize the operators $\delta(\beta), \delta(\gamma)$.

The bottom line of the discussion is that we have several inequivalent ways to soak up Bose zero-modes which differ by the operation of *picture changing*. How to compute correlation functions with $\delta(\beta), \delta(\gamma)$ inserted will be described in Sect. 10.2. There we shall provide additional details about picture changing in the path integral formalism.

Exercise 2.17 Check that the two sides of (2.300) have the same weight and charge.

2.6 Inclusion of Boundaries: Non-orientable Surfaces

Up to now we have discussed CFT on an oriented surface without boundaries. To describe processes involving open strings we need to include boundaries and impose appropriate boundary conditions consistent with conformal invariance. For non-oriented strings we need also to consider CFT on non-oriented world-sheets.

Boundaries

We consider a CFT on the infinite strip

$$\text{st} \equiv \{w \in \mathbb{C}: 0 \leq \text{Re } w \leq \pi\}. \quad (2.301)$$

This world-sheet describes the free propagation of an open oriented string.

The conformal transformation

$$w \rightarrow z = -\exp(-iw) \quad (2.302)$$

maps the strip st to the upper half-plane

$$\mathcal{H} = \{z \in \mathbb{C}: \text{Im } z > 0\} \quad (2.303)$$

which is the Hermitian symmetric space $SL(2, \mathbb{R})/U(1)$. The Cayley transform [47]

$$z \rightarrow z' = \frac{z - i}{z + i} \quad (2.304)$$

yields a biholomorphic equivalence⁵⁹ between $\mathcal{H}$ and the unit disk

$$\mathcal{D} = \{z' \in \mathbb{C}: |z'| < 1\}. \quad (2.305)$$

⁵⁹ More generally, recall that by uniformization (BOX 1.9) all simply connected oriented surfaces which are not $\mathbb{P}^1$ or $\mathbb{C}$ are biholomorphically equivalent: hence the strip, the upper half-plane, the disk, and *any other* simply connected domain $\subsetneq \mathbb{C}$ are conformally equivalent.

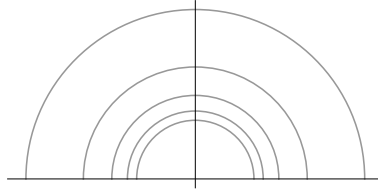


Fig. 2.1 Radial quantization for a CFT defined on a strip. The strip is conformally mapped to the upper half-plane $\mathcal{H}$, while the lines of constant Euclidean time $t \equiv \text{Im } w$ in the strip are mapped into half-circles in the upper half-plane centered at the origin

It is more convenient to work with $\mathcal{H}$ whose boundary is the real axis $\mathbb{R}$ plus the point $i\infty$. We write $z = x + iy$. We note that the equal-time lines on the strip,

$$\text{Im } w = t \equiv \text{const.} \quad (2.306)$$

are mapped in $\mathcal{H}$ into the semi-circles centered at the origin of equation

$$|z| = e^t; \quad (2.307)$$

see Fig. 2.1. Again, the asymptotic region $t \rightarrow -\infty$ corresponds to the origin in the closed upper half-plane $\overline{\mathcal{H}} \equiv \mathcal{H} \cup \mathbb{R} \cup \{i\infty\}$.

Conformal invariance requires that along the boundary $\partial\mathcal{H} \equiv \mathbb{R} \cup \{i\infty\}$ the energy–momentum tensor satisfies the boundary condition

$$T_{yx}(z) \Big|_{z \in \mathbb{R}} = 0. \quad (2.308)$$

Indeed $\partial_x \in \mathfrak{sl}(2, \mathbb{R})$ is a good Killing vector on $\mathcal{H}$ whose flow preserves the boundary, and therefore the current T_{ax} , which generates the corresponding conformal motion (isometry), should be conserved. In the bulk this requires $D^a T_{ax} = 0$; in addition the flow of the current through the boundary must vanish: this condition gives Eq. (2.308) or, in the complex notation,

$$T(z) = \tilde{T}(\bar{z}) \quad \text{for } z = \bar{z}. \quad (2.309)$$

Note 2.6 The boundary condition (2.309) is sufficient to guarantee conformal invariance in the presence of a boundary for all world-sheet Σ . Its general form is

$$T_{ab} n^a t^b = 0 \quad (2.310)$$

where t^a (resp. n^a) is a tangent vector (resp. normal vector) to the boundary.

In a Lagrangian field theory the energy–momentum tensor is constructed out of the fundamental fields, e.g. Eq. (2.205) for the free scalar theory. Therefore we

have to impose boundary conditions (b.c.) on the fundamental fields which *imply* the boundary condition (2.309) on the energy–momentum tensor. By classifying all such fundamental field b.c. we get all boundary conditions consistent with conformal invariance in the given Lagrangian model.

At this point it is convenient to use a theorem in complex analysis, the *Schwarz reflection principle* [48], which in the present context is usually called the *doubling trick*. We extend the energy–momentum tensor $T(z)$ from $\mathcal{H}$ to the whole $\mathbb{C}$ by *declaring* that its value in the lower half-plane is

$$T(x - iy) \stackrel{\text{def}}{=} \tilde{T}(x - iy) \equiv \tilde{T}(\bar{z}), \quad \text{for } y > 0, \text{ so } z \equiv x + iy \in \mathcal{H}. \quad (2.311)$$

In this way $T(z)$ is holomorphic in both half-planes, and continuous across the real axis by the boundary condition (2.309). The Schwarz reflection theorem [48] guarantees that $T(z)$ is actually holomorphic also along the real axis, so $T(z)$ is holomorphic *everywhere* in $\mathbb{C}$. We are thus reduced to the situation already considered before in the absence of boundaries, except that now we have *only one set* of Virasoro generators since $\tilde{T}(\bar{z})$ is no longer an independent operator:

$$L_m = \oint_C \frac{dz}{2\pi i} z^{m+1} T(z) = \int_{\substack{\text{upper} \\ \text{half-circle}}} \frac{dz}{2\pi i} z^{m+1} T(z) + \int_{\substack{\text{lower} \\ \text{half-circle}}} \frac{d\bar{z}}{-2\pi i} \bar{z}^{m+1} \tilde{T}(\bar{z}). \quad (2.312)$$

These modes satisfy the usual Virasoro algebra

$$[L_m, L_n] = (m - n) L_{m+n} + \frac{c}{12} (m^3 - m) \delta_{m+n}. \quad (2.313)$$

The subalgebra generated by $L_{\pm 1}$, L_0 generates the $SL(2, \mathbb{R})$ automorphism group of $\mathcal{H}$ acting by real Möbius transformations

$$z \rightarrow \frac{az + b}{cz + d}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{R}). \quad (2.314)$$

Operator Insertions As it is obvious from its very definition, the *doubling trick* is just the *image method* of classical 2d electrostatics. Therefore only reflection-invariant operator insertions on the doubled Σ make sense.

There are two distinct cases. First we may have local operators inserted in the bulk of the upper half-plane $\mathcal{H}$ —then, when applying the doubling trick, we have to introduce their images in the lower half-plane. A n -point function of bulk operators in the upper half-plane $\mathcal{H}$ becomes a $2n$ -point function in $\mathbb{C}$. For instance, the one-point function gets mapped to the 2-point one, so that the bulk operators' tadpoles may not vanish when $h = \tilde{h}$. In this case

$$\left\langle \phi_i(z, \bar{z}) \right\rangle_{\substack{\text{upper} \\ \text{half-plane}}} = \frac{\text{const}}{(z - \bar{z})^{2h}}. \quad (2.315)$$

The actual constant depends on the boundary condition satisfied by the operator ϕ_i along the real axis.

The second situation is an operator insertion on the real axis, which does not require a distinct image. Using the symmetry $x \rightarrow x + \text{const}$, we may transport such an insertion at the origin. In view of the state/operator correspondence on $\mathbb{C}$, this insertion yields an initial state which we may evolve to some radial time, producing a state defined on a circle $|z| = \text{const}$. If the boundary conditions on the various fields/operators are consistent with conformal invariance, this state is invariant under the reflection $z \leftrightarrow \bar{z}$, so it gives a well-defined state on the half-circle, that is, a well-defined state on the spatial interval $\text{Im } w = \text{const}$ in the strip. Hence we conclude

$$\text{local operators on the boundary} \longleftrightarrow \text{states on the interval} \quad (2.316)$$

which is the state/operator correspondence relevant for the open string. The correspondence makes sense provided the boundary conditions on the various fields/operators are consistent with conformal invariance.

Non-Orientable Surfaces

A conformal structure on a non-orientable surface Σ is a complex structure on its orientable double $\check{\Sigma}$. $\check{\Sigma}$ has an orientation-reserving involutive anti-automorphism⁶⁰

$$\Omega: \check{\Sigma} \rightarrow \overline{\check{\Sigma}} \quad (2.317)$$

without fixed points.⁶¹ Again we have a *doubling trick* which means just replacing the surface Σ by its double $\check{\Sigma}$ and then keeping only states/operators even under Ω . In practice, we still use the image method for the “reflection” Ω . Often a simpler approach is possible. As explained in **BOX 1.5**, we can see a non-oriented surface R_h as a sphereless h disk closed by h cross-caps. Consider the plane $\mathbb{C}$ with a small disk centered at the origin cut out and replaced by a cross-cap. Clearly in radial quantization this corresponds to a cross-cap state $|C\rangle$ in the radial Hilbert space $\mathbf{H}_{S^1}$ of the CFT. For instance, we may write the amplitude on $\Sigma = \mathbb{RP}^2$ ($h = 1$) as

$$\langle 0 | O_{i_1}(z_1) \cdots O_{i_s}(z_s) | C \rangle = \langle O_{i_1}(z_1) \cdots O_{i_s}(z_s) C(0) \rangle_{S^2} \quad (2.318)$$

where $C(z)$ is the local operator which corresponds to the state $|C\rangle$.

⁶⁰ An *anti-automorphism* is a diffeomorphism which maps the complex structure in its conjugate.

⁶¹ This situation should be contrasted with the case of a boundary. In that case we also have a double $\check{\Sigma}$ with an anti-automorphism Ω , but the anti-automorphism has a fixed set S which decomposes $\check{\Sigma}$ in two halves interchanged by the anti-automorphism: $\check{\Sigma} \setminus S = \Sigma_+ \sqcup \Sigma_-$ and $\Omega: \Sigma_{\pm} \leftrightarrow \Sigma_{\mp}$. Then we may identify the quotient space $\check{\Sigma}/\Omega$ with one half Σ_+ , whose boundary is the fixed set S .

2.7 KaČ–Moody and Current Algebras

We consider 2d CFT with a continuous internal chiral symmetry $G \times \tilde{G}$ where the currents of G (resp. $\tilde{G}$) are chiral left-moving (right-moving). We are interested in the way the symmetry is realized at the quantum level and in its interrelations with conformal symmetry. We also assume our CFT to be *unitary*.

Abelian Case

We already studied the case of G an Abelian group, say $G = U(1)$, when discussing the $U(1)$ current $j(z)$ associated with the free b, c . In the absence of anomalies (i.e. when $Q = 0$) $j(z)$ was a primary operator of weights $(1, 0)$ and we had the two OPEs

$$j(z) j(w) \sim \frac{1}{(z-w)^2}, \quad T(z) j(w) \sim \frac{j(w)}{(z-w)^2} + \frac{\partial j(w)}{z-w}, \quad (2.319)$$

the current algebra energy–momentum tensor $T^j(z)$

$$T^j(z) = \frac{1}{2} : j(z)^2 : , \quad (2.320)$$

and the bosonization of the current $j(z)$ in terms of a free chiral scalar field $\phi(z)$

$$j(z) = \partial\phi(z), \quad \phi(z)\phi(w) \sim \log(z-w). \quad (2.321)$$

We now consider the case in which G is a non-Abelian Lie group. The following formulae reduce to the above ones when G is Abelian.

2.7.1 KaČ–Moody Algebras

We identify S^1 with the unit circle in the complex plane

$$S^1 \equiv \{z \in \mathbb{C} : |z| = 1\}. \quad (2.322)$$

Loop Algebras

Let G be a finite-dimensional, compact, connected Lie group and $\mathfrak{g} = \mathfrak{Lie}(G)$ its Lie algebra. The *loop group* $\mathcal{G}$ of G is the group of maps from the circle S^1 to G

$$\mathcal{G} \stackrel{\text{def}}{=} \text{Maps}(S^1, G). \quad (2.323)$$

The elements of $\mathcal{G}$ are maps $z \mapsto \gamma(z) \in G$ and product is pointwise multiplication

$$\gamma_1 \cdot \gamma_2(z) = \gamma_1(z)\gamma_2(z). \quad (2.324)$$

$\mathcal{G}$ is an infinite-dimensional Lie group. Its Lie algebra $\mathfrak{L}(\mathfrak{g})$ is the *loop algebra of type $\mathfrak{g}$* .⁶²

Let $\{T^a\} (a = 1, \dots, \dim G)$ be set generators of $\mathfrak{g}$, which we take to be Hermitian, $T^{a\dagger} = T^a$, and orthonormal with respect to some invariant product on $\mathfrak{g}$. The Lie bracket reads

$$[T^a, T^b] = i f^{ab}_c T^c. \quad (2.325)$$

The loop algebra $\mathfrak{L}(\mathfrak{g})$ is just the Lie algebra $\mathfrak{g}[z^{\pm 1}]$, namely the Lie algebra $\mathfrak{g}$ defined over the ring $\mathbb{C}[z^{\pm 1}]$ of Laurent polynomials. A basis of the loop algebra $\mathfrak{L}(\mathfrak{g})$ is then given by the generators

$$T_n^a \stackrel{\text{def}}{=} T^a z^n, \quad T_n^{a\dagger} = T_{-n}^a, \quad a = 1, \dots, \dim G, \quad n \in \mathbb{Z}, \quad (2.326)$$

with bracket

$$[T_m^a, T_n^b] = i f^{ab}_c T_{m+n}^c \quad \text{loop algebra } \mathfrak{g}[z^{\pm 1}]. \quad (2.327)$$

The “zero-modes” T_0^a generate a finite-dimensional Lie subalgebra isomorphic to $\mathfrak{g}$.

The functions on S^1 carry a natural representation of the loop algebra of S^1 ($\equiv$ the Virasoro algebra with central charge $c = 0$), given by

$$L_n = -z^{n+1} \partial_z, \quad [L_m, L_n] = (m - n) L_{m+n}, \quad (2.328)$$

so we may unify the loop algebra of S^1 and the loop algebra of G into a *single* Lie algebra with generators $\{T_m^a, L_m\}_{m \in \mathbb{Z}}$ and bracket (2.327), (2.328) together with

$$[L_m, T_n^a] = -n T_{m+n}^a. \quad (2.329)$$

Quantum 2d Conformal Current Algebras

The loop algebra (2.328) is the classical symmetry of a 2d conformal field theory; at the quantum level (if the theory is non-trivial) the loop algebra gets replaced by its central extension, the Virasoro algebra

$$[L_m, L_n] = \frac{c}{12} m(m^2 - 1) \delta_{m+n,0} + (m - n) L_{m+n}, \quad (2.330)$$

which is the algebra of quantum modes of the spin-2 chiral current $T(z)$. Likewise, the loop algebra $\mathfrak{g}[z^{\pm 1}]$ is just the classical version of the quantum current algebra of G , which is the algebra of quantum modes of the spin-1 (Noether) chiral currents $J^a(z)$ associated with the generators of the chiral (left-moving) symmetry G . We

⁶² More generally, given a compact manifold M we may consider the group $\mathcal{G}_M$ of maps $M \rightarrow G$. If M is a torus $(S^1)^k$, the Lie algebra of $\mathcal{G}_M$ is called a *torus loop algebra* and its central extensions are called *Extended Affine Lie Algebras* (EALA) of type $\mathfrak{g}$ and nullity k [49, 50]. For $k > 1$ the EALA are not Kač–Moody algebras since their Cartan matrices do not have non-positive off-diagonal entries. Rather $k > 1$ EALA are Slodowy–GIM Lie algebras [51].

expect that the appropriate quantum (conformal) current algebra is also a central extension of the classical loop algebra of G , Eq. (2.327).

There is a mathematical classification of all possible central extensions modulo the trivial ones [52]: one shows that for G simple there is a one-parameter family of such extensions

$$[T_m^a, T_n^b] = m \hat{k} \delta^{ab} \delta_{m+n} + i f^{ab}_c T_{m+n}^c, \quad m, n \in \mathbb{Z} \quad (2.331)$$

where $\hat{k}$ is a central element called the *central charge*. In each irreducible representation $\hat{k}$ is just a fixed number. The numerical value of $\hat{k}$ depends on the normalization of the generators of the algebra; its convention-independent counterpart is

$$k = \frac{2 \hat{k}}{\|\theta\|^2}, \quad (2.332)$$

where $\|\theta\|^2$ is the squared-length of the highest root θ of $\mathfrak{g}$. k is called the *level* of the current algebra.

The infinite-dimensional Lie algebra (2.331) is usually called⁶³ the *Kač–Moody algebra* of type $\mathfrak{g}$, written as $\tilde{\mathfrak{g}}$ or G_k when we wish to emphasize the level k . Its technical name is *the untwisted affine Lie algebra of type $\mathfrak{g}$ and level k* [52, 53].

The Lie algebra (2.331) is isomorphic to the algebra of (modes of) chiral spin-1 currents in a CFT with internal symmetry G . For each generator T^a we have a left-moving chiral current $J^a(z)$ which is a primary⁶⁴ operator of weight $(h, \tilde{h}) = (1, 0)$ which—being holomorphic—is automatically conserved

$$\bar{\partial} J^a(z) = 0. \quad (2.333)$$

Their mode expansions are

$$J^a(z) = \sum_{n \in \mathbb{Z}} \frac{J_n^a}{z^{n+1}}, \quad J_n^a = \oint \frac{dz}{2\pi i} z^n J^a(z), \quad (2.334)$$

with commutators

$$[J_m^a, J_n^b] = i f^{ab}_c T_{m+n}^c + m \hat{k} \delta^{ab} \delta_{m+n}, \quad [L_m, J_n^a] = -n J_{m+n}^a, \quad (2.335)$$

and the second equation states that $J^a(z)$ is a Virasoro primary with $h = 1$; cf. (2.114).

⁶³ In the physicists' jargon. In mathematicians' parlance, the Kač–Moody algebras form a more general class of Lie algebras [52].

⁶⁴ Recall from the discussion in Sect. 2.5.2 that if a spin-1 chiral current $j(z)$ is not Virasoro primary, then the corresponding symmetry is anomalous when quantized on a surface Σ with non-zero scalar curvature R . In 2d the *non-Abelian* chiral currents cannot have such “gravitational” anomalies, so the corresponding currents are necessarily primaries. Alternatively apply Exercise 2.6.

The commutators (2.335) are equivalent to the OPEs

$$J^a(z) J^b(w) = \frac{\hat{k} \delta^{ab}}{(z-w)^2} + i f^{ab}{}_c \frac{J^c(w)}{(z-w)} + \text{reg.} \quad (2.336)$$

$$T(z) J^a(w) = \frac{J^a(w)}{(z-w)^2} + \frac{\partial J^a(w)}{(z-w)} + \text{reg.} \quad (2.337)$$

These two equations can be easily derived directly from the fundamental principles. The second one just follows from the fact that a conserved (left-)chiral current $J(z)$ is a holomorphic 1-form. First one: by locality and dimension considerations, the OPE of two currents has, at most, a double pole proportional to the identity operator and a simple pole whose residue is a local operator of weights $(1, 0)$, hence a linear combination of the currents themselves.⁶⁵ The coefficients of the singular part of the OPE must be G -invariant tensors, the first one symmetric in the indices of the currents, and the second one antisymmetric. For G a *simple* Lie group, this implies the coefficient of $(z-w)^{-2}$ is proportional to δ^{ab} , while the residue of $(z-w)^{-1}$ is proportional to $i f^{ab}{}_c J^c$. Fixing the normalization of the currents so that the single pole has the standard form, we are left with a single free parameter $\hat{k}$.

In an unitary CFT the level is non-negative $k \geq 0$. Indeed, since the currents are Virasoro primaries, by Eq. (2.130) $J_n^a|0\rangle = 0$ for $n \geq 0$ so (not summed over a !)

$$0 \leq \|J_{-n}^a|0\rangle\|^2 = \langle 0|J_n^a J_{-n}^a|0\rangle = \langle 0|[J_n^a, J_{-n}^a]|0\rangle = n \hat{k}, \quad \text{for } n \geq 0. \quad (2.338)$$

Claim 2.3 For G *simple* the level k must be an integer.

Proof We write H_i for the Cartan generators of G . Let α be a root of $\mathfrak{g}$ and E^α a generator in the α -root space normalized so that [54, 55]

$$J^\pm = E^{\pm\alpha}, \quad J^3 = \frac{(\alpha, H)}{(\alpha, \alpha)} \quad (2.339)$$

generates the $\mathfrak{sl}(2)$ algebra

$$[J^3, J^\pm] = \pm J^\pm, \quad [J^+, J^-] = 2J^3. \quad (2.340)$$

We consider the three generators in the algebra (2.331)

$$J^\pm = E_{\pm 1}^{\pm\alpha}, \quad J^3 = \frac{\alpha \cdot H_0 + \hat{k}}{\alpha^2}. \quad (2.341)$$

They also generate a $\mathfrak{sl}(2)$ algebra (2.340). In all finite-dimensional representations of $\mathfrak{sl}(2)$, $2J^3$ is an integer. $2\alpha \cdot H_0/\alpha^2$ is an integer (it is $2J^3$ for the $\mathfrak{sl}(2)$ algebra (2.339)) and so $2\hat{k}/\alpha^2 \in \mathbb{N}$. $\square$

The Chiral Algebra $\mathcal{A}$

Recall that in Sect. 2.3.6 we introduced the concept of a general chiral algebra $\mathcal{A}$ containing Virasoro as a subalgebra. We see that the semi-direct sum of the Virasoro and Kač–Moody algebras is an example of such generalized chiral algebra generated

⁶⁵ In a unitary CFT an operator of weight $(1, 0)$ is automatically primary; cf. Exercise 2.6.

by one spin-2 holomorphic current $T(z)$ and a number $\dim G$ of spin-1 holomorphic currents $J^a(z)$. It is natural (and convenient) to study the representation theory of the full algebra $\mathcal{A}$ rather than the one of the Kač–Moody algebra *per se*.

We define the primary fields $\phi_i(z)$ of $\mathcal{A}$ by the two OPEs

$$T(z) \phi_i(w) = \frac{h \phi_i(w)}{(z-w)^2} + \frac{\partial \phi_i(w)}{(z-w)} + \dots \quad (2.342)$$

$$J^a(z) \phi_i(w) = \frac{(T^a)_i^j \phi_j(w)}{z-w} + \dots \quad (2.343)$$

where $(T^a)_i^j$ are the matrices representing the finite-dimensional Lie algebra $\mathfrak{g}$ in some *right*⁶⁶ representation. Note that while the currents $J^a(z)$ are Virasoro primaries, they are not primary for the full $\mathcal{A}$, but rather descendants of the identity.

As for all chiral algebras, the *highest weight states* (a.k.a. primary states) are obtained by applying the CFT state-operator correspondence to the primary operators

$$|\phi_i\rangle = \phi_i(0)|0\rangle, \quad (2.344)$$

here i is the index of some representation of G and $\phi_i(z)$ is primary.

Exercise 2.18 Show that the $SL(2, \mathbb{C})$ -invariant state $|0\rangle$ is also G -invariant.

The primary states of $\mathcal{A}$ then satisfy

$$L_n |\phi_i\rangle = 0 \quad n > 0 \quad L_0 |\phi_i\rangle = h_i |\phi_i\rangle \quad (2.345)$$

$$J_n^a |\phi_i\rangle = 0 \quad n > 0 \quad J_0^a |\phi_i\rangle = (T^a)_i^j |\phi_j\rangle. \quad (2.346)$$

The descendant states have the form

$$L_{-k_1} \cdots L_{-k_m} J_{-l_1}^{a_1} \cdots J_{-l_n}^{a_n} |\phi_i\rangle, \quad k_i, l_i > 0, \quad (2.347)$$

with conformal weight

$$h_i + \sum_{i=1}^m k_i + \sum_{j=1}^n l_j. \quad (2.348)$$

The states (2.347) form the current algebra family $[\phi_i]_{\mathcal{A}}$.

2.7.2 The Sugawara Construction

There is an explicit construction of the Virasoro generators L_m in terms of the current modes J_n^a which is the generalization of the current algebra energy–momentum tensor

⁶⁶ That is, the transpose matrices $(T^a)^t$ yield an ordinary (i.e. left) representation.

$T^j(z)$ of Sect. 2.5.4 to the non-Abelian case. Without loss of generality we may (and do) assume the Lie algebra $\mathfrak{g}$ to be simple.

We define the *Sugawara energy–momentum tensor* $T(z)^{\text{sug}}$ as

$$T(z)^{\text{sug}} \stackrel{\text{def}}{=} \frac{1}{2\hat{k} + C_2} \sum_a : J^a(z) J^a(z) : \equiv \sum_{n \in \mathbb{Z}} \frac{L_n^{\text{sug}}}{z^{n+2}} \quad (2.349)$$

$$\text{with } L_n^{\text{sug}} = \frac{1}{2\hat{k} + C_2} \sum_a \sum_{m \in \mathbb{Z}} : J_m^a J_{n-m}^a : , \quad (2.350)$$

where C_2 is the quadratic Casimir invariant of the adjoint representation defined by

$$- f^{ac}{}_d f^{bd}{}_c = C_2 \delta^{ab}, \quad (2.351)$$

whose value depends on the normalization of the generators. We can relate C_2 to a normalization independent quantity, namely the *dual Coxeter number* $h_{\mathfrak{g}}^{\vee}$

$$h_{\mathfrak{g}}^{\vee} = \frac{C_2}{\|\theta\|^2} \in \mathbb{N} \quad (2.352)$$

with $\|\theta\|^2$ the square-length of a long root (see **BOX 2.8** for Lie theory background).

The normal order in (2.349) is defined in the usual way. In terms of modes

$$: J_n^a J_m^b : = \begin{cases} J_n^a J_m^b & n \leq -1 \\ J_m^b J_n^a & n \geq 0. \end{cases} \quad (2.353)$$

In **BOX 2.9** it is shown that the current–current OPE (2.336) implies the following OPEs for the Sugawara energy–momentum tensor

$$T(z)^{\text{sug}} J^a(w) = \frac{J^a(w)}{(z-w)^2} + \frac{\partial J^a(w)}{(z-w)} + \text{reg.} \quad (2.354)$$

$$T(z)^{\text{sug}} T(w)^{\text{sug}} = \frac{c^{\text{sug}}}{2(z-w)^4} + \frac{2 T(w)^{\text{sug}}}{(z-w)^2} + \frac{\partial T(w)^{\text{sug}}}{z-w} + \text{reg.} \quad (2.355)$$

where the *Sugawara central charge* c^{sug} is

$$c^{\text{sug}} \stackrel{\text{def}}{=} \frac{2\hat{k} \dim G}{2\hat{k} + C_2} = \frac{k \dim G}{k + h_{\mathfrak{g}}^{\vee}}. \quad (2.356)$$

We see that the Sugawara energy–momentum tensor $T(z)^{\text{sug}}$ satisfies the Virasoro OPE with the central charge (2.356).

BOX 2.8 - Coxeter number and dual Coxeter number

A Coxeter element $C_{\mathfrak{g}} \in \text{Weyl}(\mathfrak{g})$ of the simple Lie algebra $\mathfrak{g}$ is the product of all simple reflections for some choice of Weyl chamber and ordering of the simple roots. All Coxeter elements are conjugate in $\text{Weyl}(\mathfrak{g})$, so its order, called the *Coxeter number* $h_{\mathfrak{g}}$ of $\mathfrak{g}$, is well-defined independently of choices [55–57]. One has the Coxeter identity

$$\dim \mathfrak{g} = (h_{\mathfrak{g}} + 1)r_{\mathfrak{g}}$$

where $r_{\mathfrak{g}}$ is the rank of $\mathfrak{g}$. The standard definition of the Coxeter number $h_{\mathfrak{g}}$ refers to the expansion of the highest root θ in terms of the simple roots α_i . One has [55, 56]

$$\theta = \sum_{i=1}^{r_{\mathfrak{g}}} m_i \alpha_i \quad \text{for } m_i \text{ positive integers,} \quad h_{\mathfrak{g}} \stackrel{\text{def}}{=} 1 + \sum_i m_i.$$

The *dual Coxeter number* has a similar expression in terms of the maximal coroot $\theta^{\vee}$ [55]

$$h_{\mathfrak{g}}^{\vee} = 1 + \sum_i p_i, \quad \theta^{\vee} = \sum_i p_i \alpha_i^{\vee}.$$

In particular, for simply laced Lie algebras (types A_n, D_n, E_6, E_7, E_8) one has $h_{\mathfrak{g}} \equiv h_{\mathfrak{g}}^{\vee}$. For non-simply laced Lie algebras $h_{\mathfrak{g}}^{\vee} < h_{\mathfrak{g}}$. Indeed, since the maximal root is always a long root, we have $\langle \alpha_i, \alpha_i \rangle = \langle \theta, \theta \rangle / d_i$, with d_i integers ≥ 1 (cf. Figs. 2.2 and 2.3a–c)

$$\alpha_i^{\vee} = \frac{2}{\langle \theta, \theta \rangle} d_i \alpha_i \quad \text{and} \quad \theta^{\vee} = \frac{2\theta}{\langle \theta, \theta \rangle} = \sum_i \frac{m_i}{d_i} \alpha_i^{\vee} \quad \Rightarrow \quad h_{\mathfrak{g}}^{\vee} = 1 + \sum_i \frac{m_i}{d_i},$$

which (using Figs. 2.2 and 2.3a–c) yields the following table for

$\mathfrak{g}$	A_n	B_n	C_n	D_n	E_6	E_7	E_8	F_4	G_2
$h_{\mathfrak{g}}$	$n+1$	$2n$	$2n$	$2n-2$	12	18	30	12	6
$h_{\mathfrak{g}}^{\vee}$	$n+1$	$2n-1$	$n+1$	$2n-2$	12	18	30	9	4

Relation with the Quadratic Casimir Let $\langle \cdot, \cdot \rangle$ be an invariant quadratic form on $\mathfrak{g}$ we use to define the quadratic Casimir. Then it is easy to see that the value of the quadratic Casimir on the irreducible representation of highest weight λ is

$$C_2 = \langle \lambda, \lambda + 2\rho \rangle$$

where ρ is half the sum of the positive roots which is also the sum of all fundamental weights ω_i defined by $\langle \alpha_i^{\vee}, \omega_j \rangle = \delta_{ij}$. To get the Casimir of the adjoint representation, we have to specialize this formula to the case $\lambda = \theta$. Using $\theta^{\vee} = 2\theta / \langle \theta, \theta \rangle$, we get

$$C_2 = \langle \theta, \theta \rangle + \langle \theta, \theta \rangle \cdot \langle \theta^{\vee}, \rho \rangle = \langle \theta, \theta \rangle \left(1 + \sum_{i,j} p_i \langle \alpha_i^{\vee}, \omega_j \rangle \right) \equiv \|\theta\|^2 \cdot h_{\mathfrak{g}}^{\vee}.$$

BOX 2.9 - OPEs for the Sugawara energy–momentum tensor

In this **BOX** we prove Eqs. (2.354), (2.355). These OPEs are a bit subtle since the *non-Abelian* current algebra corresponds to an *interacting* CFT rather than a free one as in the Abelian case.

We start from the operator product identity (not summed over a , summed over c)

$$\begin{aligned} J^a(z) J^a(u) J^b(w) &= \frac{\hat{k}}{(z-w)^2} J^b(u) + i \frac{f^{abc}}{z-w} J^c(z) J^a(u) + \\ &+ \frac{\hat{k}}{(u-w)^2} J^b(z) + i \frac{f^{abc}}{u-w} J^c(u) J^a(z) + \text{holomorphic in } w \end{aligned}$$

which is proven by comparing the singularities of the two sides as $w \rightarrow z$ and $w \rightarrow u$. Now, taking into account that f^{abc} is totally antisymmetric, we have (now summing over a !)

$$\begin{aligned} i \frac{f^{abc}}{z-w} J^c(z) J^a(u) + i \frac{f^{abc}}{u-w} J^c(u) J^a(z) &\sim \frac{f^{abc} f^{adc}}{z-u} \left(\frac{1}{z-w} - \frac{1}{u-w} \right) J^d(z) = \\ &= \frac{C_2}{(z-w)(u-w)} \left(J^b(w) + (z-w) \partial J^b(w) + O((z-w)^2) \right) \end{aligned}$$

and then

$$\lim_{u \rightarrow z} : J^a(z) J^a(u) : J^b(w) \sim (2\hat{k} + C_2) \left(\frac{J^b(w)}{(z-w)^2} + \frac{\partial J^b(w)}{z-w} \right),$$

so that the energy–momentum tensor $T(z)$ and the Sugawara tensor

$$T^{\text{sug}}(z) \stackrel{\text{def}}{=} \frac{1}{2\hat{k} + C_2} : J^a J^a(z) :$$

♣

have the same (singular part of the) OPE with the currents $J^a(z)$, Eq. (2.354). Next consider the operator product

$$\begin{aligned} J^a(z) J^a(u) T^{\text{sug}}(w) &= \frac{1}{(w-z)^2} J^a(z) J^a(u) + \frac{1}{w-z} \partial J^a(z) J^a(u) + \\ &+ \frac{1}{(w-u)^2} J^a(u) J^a(z) + \frac{1}{w-u} \partial J^a(u) J^a(z) + \text{holomorphic in } w. \end{aligned}$$

In the RHS the term proportional to the identity operator is

$$\begin{aligned} \hat{k} \dim G \left[\left(\frac{1}{(w-z)^2} + \frac{1}{(w-u)^2} \right) \frac{1}{(z-u)^2} + \frac{1}{w-z} \partial_z \frac{1}{(z-u)^2} + \frac{1}{w-u} \partial_u \frac{1}{(u-z)^2} \right] \\ = \hat{k} \dim G \sum_{k \geq 0} \frac{(k+1)(z-u)^k}{(z-w)^{k+4}} \end{aligned}$$

$$\text{that is, } : J^a(z) J^a(z) : T^{\text{sug}}(w) = \frac{\hat{k} \dim G}{(z-w)^4} + \frac{2 : J^a(w) J^a(w) :}{(z-w)^2} + \frac{\partial (: J^a(w) J^a(w) :)}{z-w} + \text{regular}$$

which, in view of the definition (♣), is equivalent to Eq. (2.355).

As in the Abelian case, we may define

$$T'(z) \stackrel{\text{def}}{=} T(z) - T(z)^{\text{sug}}. \quad (2.357)$$

Since the Sugawara energy–momentum tensor has the same singular OPE with the current $J^a(w)$ as the energy–momentum tensor $T(z)$ of the CFT, we have

$$T'(z) J^a(w) = \text{regular as } z \rightarrow w, \quad (2.358)$$

hence $T'(z) T(w)^{\text{sug}}$ is also regular. Then

$$T'(z) T'(w) \sim \frac{c'}{2(z-w)^4} + \frac{2T'(z)}{(z-w)^2} + \frac{\partial T'(w)}{z-w} \quad (2.359)$$

where

$$c' \equiv c - c^{\text{sug}} \geq 0 \quad (2.360)$$

where the inequality follows from unitarity. If the inequality is saturated, $T'(z) \equiv 0$ and the CFT consists of the G -current algebra only. In this case we can write the CFT as a 2d *Wess–Zumino–Witten model*: this is the non-Abelian version of 2d bosonization; see **Appendix 1** to this chapter. In the general situation the CFT decomposes in the Sugawara current algebra system and a decoupled “residual” CFT with central charge c' .

Primary Fields: Restriction from Unitarity

We focus on the case in which $T'(z) \equiv 0$, i.e. we assume that our (unitary) CFT consists of the current algebra CFT only. The current algebra CFT has several Lagrangian realizations (bosonic as well as fermionic); see below and **Appendix 1**.

As we saw above, a primary operator/state of the chiral algebra $\mathcal{A}$ carries a (unitary, *right*) representation R of the Lie algebra $\mathfrak{g}$

$$J_0^a |\phi_i\rangle = (T^a)_i{}^j |\phi_j\rangle \quad (\text{cf. Eq. (2.346)}). \quad (2.361)$$

Then

$$h_j |\phi_j\rangle \equiv L_0^{\text{sug}} |\phi_i\rangle = \frac{1}{2\hat{k} + C_2} (T^a)_i{}^j (T^a)_j{}^k |\phi_k\rangle = \frac{C(R)}{2\hat{k} + C_2} |\phi_i\rangle, \quad (2.362)$$

where $C(R)$ is the quadratic Casimir of the representation R . For example, for $SU(2)_k$ the Virasoro weights of a primary field of isospin j is

$$h_j = \frac{j(j+1)}{k+2}. \quad (2.363)$$

Claim 2.4 For any given level k , the primary operators ϕ_i belong to a *finite set* $\{R\}_k$ of irreducible representations of $\mathfrak{g}$.

Proof We return to the $\mathfrak{sl}(2)$ Lie subalgebra of the Kač–moody algebra defined in Eqs. (2.341). We consider a primary state with highest $\mathfrak{g}$ weight $|\lambda\rangle$; then

$$0 \leq \|E_{-1}^\alpha |\lambda\rangle\|^2 = \langle \lambda | E_1^{-\alpha} E_{-1}^\alpha | \lambda \rangle = \langle \lambda | [E_1^{-\alpha}, E_{-1}^\alpha] | \lambda \rangle = \frac{1}{\alpha^2} (2\hat{k} - \alpha \cdot \lambda). \quad (2.364)$$

The inequality is sharper when we take α to be the highest root θ

$$|(\theta^\vee, \lambda)| = \frac{2}{\|\theta\|^2} |(\theta, \lambda)| \leq \frac{2\hat{k}}{\|\theta\|^2} \equiv k. \quad (2.365)$$

Let $[a_1, \dots, a_{r_{\mathfrak{g}}}]$ be the Dynkin label of a representation R of G ; by definition [55] this means that its highest weight has the form

$$\lambda = \sum_i a_i \omega_i \quad (2.366)$$

with ω_i the fundamental weights of $\mathfrak{g}$ (see **BOX 2.8**); a_i is a non-negative integer attached to the i th node of the Dynkin graph of $\mathfrak{g}$. The fundamental representation corresponding to the i th node is the one with $a_i = 1$ and $a_j = 0$ for $j \neq i$. Then, with the notation in **BOX 2.8**, Eq. (2.365) becomes

$$\sum_i p_i a_i \leq k. \quad (2.367)$$

Applying this equation to the $SL(2, \mathbb{R})$ -vacuum $|0\rangle$ gives $k \geq 0$. □

In particular all unitary Sugawara current algebras are RCFTs.

Rule 2.1 The representations allowed for primary states/operators of the $\mathfrak{g}$ -current algebra at level $k = 1$ are the trivial one together with the fundamental ones associated with nodes of the Dynkin graph with Coxeter *co*-label $p_i = 1$.

For simply laced Lie algebras $\mathfrak{g}$, the nodes with $p_i = 1$ are the *extension nodes*⁶⁷ of the Dynkin graph: their number is $|Z(G)| - 1$ where $Z(G)$ is the center of the simply connected Lie group with Lie algebra $\mathfrak{g}$. The non-simply laced Dynkin graphs are the quotient of a simply laced one by an automorphism group $\mathfrak{G}$ (“diagram folding”); see Fig. 2.2. The nodes with $p_i = 1$ are the ones representing $\mathfrak{G}$ -orbits of extension nodes. d_i is the number of nodes in the orbit, and $m_i = d_i p_i$.

2.7.3 ★ Knizhnik–Zamolodchikov Equation

We wish to compute the correlators of operators in the current algebra CFT with group G and level k (so that $T(z)$ is the Sugawara energy–momentum tensor). It is enough to compute the correlators between primary operators $O_i(z)$, since all the other correlators are obtained by acting with the chiral symmetry $\mathcal{A}$ on the primary operator functions. Recall the OPE

$$J^a(z) O_i(z_i) \sim \frac{T_{(i)}^a O_i(z_i)}{z - z_i} \quad (2.368)$$

⁶⁷ The extension nodes in a simply laced Dynkin graph are the ones with $m_i = 1$. The name follows for their interpretation in the extended (i.e. affine) Dynkin graph. See **Appendix 2**.

is a holomorphic function on $\mathbb{C}$ vanishing at infinity, hence zero by the Liouville theorem. We have

$$J^a(z) O_1(z_1) = \sum_{k \in \mathbb{Z}} (z - z_1)^{k-1} J_{-k}^a \cdot O_1(z_1) \quad (2.373)$$

hence, from the equality of the two terms in (2.372),

$$\begin{aligned} \left\langle J_{-k}^a \cdot O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} &= \oint_{C_{z_1}} \frac{dz}{2\pi i} (z - z_1)^{-k} \left\langle J^a(z) O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} = \\ &= - \sum_{i=2}^n \frac{T_{(i)}^a}{(z_i - z_1)^k} \left\langle O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} \end{aligned} \quad (2.374)$$

and, using Eq. (2.369),

$$\begin{aligned} \partial_{z_1} \left\langle O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} &= \left\langle L_{-1} \cdot O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} = \\ &= \frac{2}{2\hat{k} + C_2} \sum_a T_{(1)}^a \left\langle J_{-1}^a \cdot O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} = \\ &= - \frac{2}{2\hat{k} + C_2} \sum_a \sum_{i=2}^n \frac{J_{(1)}^a J_{(i)}^a}{z_i - z_1} \left\langle O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} \end{aligned} \quad (2.375)$$

where the matrix $T_{(i)}^a$ acts on the i th field $O_i(z_i)$. We define the first-order differential operators

$$\mathcal{D}_i \stackrel{\text{def}}{=} \frac{\partial}{\partial z_i} - \frac{2}{2\hat{k} + C_2} \sum_a \sum_{j \neq i} \frac{T_{(i)}^a T_{(j)}^a}{z_i - z_j}, \quad i = 1, 2, \dots, n, \quad (2.376)$$

which act on functions Ψ defined on

$$C_n \stackrel{\text{def}}{=} \left\{ (z_1, \dots, z_n) \in \mathbb{C}^n : z_i \neq z_j \text{ for } i \neq j \right\}, \quad (2.377)$$

which take values in the vector space

$$R_{(1)} \otimes R_{(2)} \otimes \cdots \otimes R_{(n)} \equiv \mathbf{R} \quad (2.378)$$

($T_{(i)}^a$ acts on the i th tensor factor of $\mathbf{R}$). Then the correlator

$$\Psi \equiv \left\langle O_1(z_1) \cdots O_n(z_n) \right\rangle_{S^2} \quad (2.379)$$

is a solution to the system of linear PDEs

$$\mathcal{D}_i \Psi = 0, \quad i = 1, 2, \dots, n. \quad (2.380)$$

Equation (2.380) is called the *Knizhnik–Zamolodchikov equation* [58].

We may see $\mathcal{D}_i$ as a covariant derivative acting on a vector bundle over C_n with typical fiber $\mathbf{R}$. The existence of a non-zero solution Ψ of Eq. (2.380) requires its curvature to vanish

$$F_{ij} \equiv [\mathcal{D}_i, \mathcal{D}_j] = 0. \quad (2.381)$$

Proposition 2.2 $F_{ij} \equiv 0$ for all simple Lie group G and all level k .

This statement follows from the theory of the *Khono connections* [59–61] and is equivalent to the classical *Yang–Baxter equation* [61, 62]. A *Khono connection* on C_n has the form

$$\mathcal{D} = d + \sum_{i < j} B_{ij} \omega_{ij}, \quad \text{where } \omega_{ij} \stackrel{\text{def}}{=} \frac{dz^i - dz^j}{z^i - z^j} \quad (2.382)$$

and the B_{ij} 's are constant matrices acting on some vector space V .

Lemma 2.1 *A Knono connection is flat, $\mathcal{D}^2 = 0$, iff the matrices B_{ij} satisfy the following equations (called the infinitesimal braid relations)*

$$[B_{ij}, B_{ik} + B_{jk}] = [B_{ij} + B_{ik}, B_{jk}] = 0 \quad \text{for } i < j < k \quad (2.383)$$

$$[B_{ij}, B_{kl}] = 0 \quad \text{for distinct } i, j, k, l. \quad (2.384)$$

If these relations are satisfied the connection $d + \lambda \sum_{i < j} B_{ij} \omega_{ij}$ is flat for all values of the parameter $\lambda \in \mathbb{C}$.

Proof (of Lemma) The closed 1-forms ω_{ij} generate a graded algebra whose only relations are the Arnold ones [63]

$$\omega_{ij} \wedge \omega_{jk} + \omega_{jk} \wedge \omega_{ki} + \omega_{ki} \wedge \omega_{ij} = 0, \quad (2.385)$$

i.e. the cohomology ring $H^*(C_n, \mathbb{C})$ is generated by the 1-forms ω_{ij} with the relations (2.385). $\square$

Proof (of Proposition) We have $B_{ij} = \lambda T_{(i)}^a T_{(j)}^a$ with $\lambda = 2(2\hat{k} + C_2)^{-1}$ and

$$2 T_{(i)}^a T_{(j)}^a + 2 T_{(j)}^a T_{(k)}^a + 2 T_{(k)}^a T_{(i)}^a = (T_{(i)}^a + T_{(j)}^a + T_{(k)}^a)^2 - T_{(i)}^a T_{(i)}^a - T_{(j)}^a T_{(j)}^a - T_{(k)}^a T_{(k)}^a \quad (2.386)$$

not summed over i, j , and k ! The second line is the quadratic Casimir of the representation $R_{(i)} \otimes R_{(j)} \otimes R_{(k)}$ minus the quadratic Casimirs of the three factor representations. Hence the matrix $B_{ij} + B_{jk} + B_{ik}$ is proportional to the identity and commutes with everything. $\square$

2.7.4 Simply Laced G at Level 1

Using the Coxeter identity (BOX 2.8) we have for the Sugawara central charge

$$c_{\mathfrak{g}} = \frac{k(h_{\mathfrak{g}} + 1)}{h_{\mathfrak{g}}^{\vee} + k} r_{\mathfrak{g}} \geq \frac{h_{\mathfrak{g}} + 1}{h_{\mathfrak{g}}^{\vee} + 1} r_{\mathfrak{g}} \geq r_{\mathfrak{g}} \quad (2.387)$$

where the first inequality is saturated iff $k = 1$ and the second one iff the Lie algebra $\mathfrak{g}$ is simply laced (i.e. iff $h_{\mathfrak{g}}^{\vee} = h_{\mathfrak{g}}$). The physical meaning of the inequality (2.387) is clear: the maximal torus of G is $U(1)^{r_{\mathfrak{g}}}$, and we can bosonize this Abelian subgroup in terms of $r_{\mathfrak{g}}$ free scalars ϕ_i which contribute to the central charge

$$c^{\text{Ab}} \equiv r_{\mathfrak{g}}. \quad (2.388)$$

We consider the Abelian current algebra energy–momentum tensor $T(z)^{\text{Ab}}$

$$T(z)^{\text{Ab}} = \frac{1}{2} \sum_i : \partial \phi_i \partial \phi_i : \quad (2.389)$$

with a “residual” energy–momentum tensor $T' = T^{\text{sg}} - T^{\text{Ab}}$ of central charge

$$c' = c^{\text{sg}} - c^{\text{Ab}} \equiv c^{\text{sg}} - r_{\mathfrak{g}} \geq 0 \quad (2.390)$$

by unitarity. Whenever the bound is saturated, $T' \equiv 0$, and the free scalars are the full current algebra CFT. This happens precisely for simply laced algebras at level 1.

We conclude that there should exist a way of constructing the current algebra of any simply laced Lie group G at level 1 in terms of $r_{\mathfrak{g}}$ (chiral) free scalars. These scalars should be compact (that is, periodic) since the Lie group G is compact.

Such a construction exists: mathematicians call it the *Frenkel–Kač–Segal construction* [64, 65]. We shall not present its abstract construction.⁶⁹ We shall focus instead on the example most relevant for superstring theory, that is, the case of the Lie algebra D_n (a.k.a. $\mathfrak{so}(2n)$) of rank n . The extension to general simply laced Lie algebras is then pretty obvious and left to the reader. However, in order to put things in the proper *superstring* perspective, we postpone the discussion to Sect. 2.9 after having introduced the superconformal field theories (SCFT).

Note 2.7 The bosonization of non-simply laced Lie algebra is reduced to the simply laced case by the diagram folding construction of Fig. 2.2; see [66].

2.7.5 Fermionic Realization of the Current Algebra

Before going to supersymmetry, let us make another fundamental remark. Consider a Weyl fermion with weights $(h, \bar{h}) = (\frac{1}{2}, 0)$ and free Dirac action

$$\frac{1}{2\pi} \int \bar{\chi} \bar{\partial} \chi. \quad (2.391)$$

This system is the same as a b, c free CFT with $\lambda = \frac{1}{2}$ *except* for the Hermitian condition that in the present case reads $\chi^\dagger = \bar{\chi}$:⁷⁰ one has

$$\chi = (\lambda_1 + i\lambda_2)/\sqrt{2}, \quad \bar{\chi} = (\lambda_1 - i\lambda_2)/\sqrt{2} \quad (2.392)$$

with λ_1, λ_2 Majorana–Weyl spinors. This CFT, consisting of *two* free Majorana–Weyl fermions, has central charge

$$c = 2 \times \frac{1}{2} \equiv 1, \quad (2.393)$$

and a non-anomalous $U(1)$ current $\bar{\chi}\chi$ which is bosonized in the form $\partial\phi$, and is equivalent to (the left-moving part of) a compact free scalar (with appropriate periodicity).

⁶⁹ The interested reader is referred to the original papers [64, 65].

⁷⁰ The difference in the Hermitian structure introduces some extra factor i in some of the bosonization formulae, which are otherwise the same ones as in Sect. 2.5.

The equivalence of r_g chiral scalars with $2r_g$ Majorana–Weyl fermions suggests that we may as well construct a level-1 simply laced current algebra in terms of $2r_g$ *free* Majorana–Weyl fermions. This is indeed the case.

The fermionic construction is the most natural one from the superstring perspective. For this reason, we shall mostly adopt the *free-fermion* approach when working with level-1 simply laced 2d current algebras. To motivate its physical relevance, consider the matter part of the world-sheet theory of the superstring, Eq. (1.140). It contains a fermionic free theory of the form (we write only the left-movers)

$$\int \psi^\mu \bar{\partial} \psi_\mu \quad (2.394)$$

where ψ^μ are 2d Majorana–Weyl fermions and $\mu = 0, 1, \dots, d-1$ is a target-space vector index. After target-space Wick rotation, and assuming $d \equiv 2n$ to be *even*⁷¹ μ is a vector index of $SO(2n)$. The global $SO(2n)$ symmetry of the (Wick rotated) action (2.394) gets promoted to a chiral current algebra. The central charge of $2n$ Majorana–Weyl fermions is $n \equiv$ the rank of $\mathfrak{so}(2n)$, saturating the unitarity bound (2.387), so the current algebra has level 1 and is the *full* CFT.

In superstring theory there will be further subtleties (see next chapter), but the study of the $SO(2n)$ current algebra is essential for all these developments. In this book we shall use the fermionic construction of the current algebra again and again.

2.8 (1, 1) Superconformal Algebra

2d Superconformal Field Theories

A (Euclidean) 2d superconformal theory (SCFT) of type (p, q) is a conformal theory which contains p (resp. q) left-moving (resp. right-moving) fermionic supercurrents

$$T_F(z)^a \quad a = 1, 2, \dots, p \quad \text{and} \quad \tilde{T}_F(\bar{z})^{\tilde{a}} \quad \tilde{a} = 1, 2, \dots, q, \quad (2.395)$$

which are primary conformal fields of weights $(3/2, 0)$ and $(0, 3/2)$, respectively. The currents $T_F(z)^a$, $\tilde{T}_F(\bar{z})^{\tilde{a}}$, being conserved and fermionic, generate (conformal) supersymmetries. In fact, being chiral, i.e. (anti-)holomorphic, in flat space $\mathbb{R}^2$ they generate a *superconformal symmetry* of type (p, q) , meaning that the Poincaré supercharges are p left-handed Majorana–Weyl spinors and q right-handed ones.

In this section, we consider only (Euclidean) (1,1) SCFTs (a.k.a. $\mathcal{N} = 1$ SCFTs) which we quantize on a Riemann surface Σ . *Extended* superconformal algebras will be classified and studied in Sect. 2.10.

The supercurrent $T_F(z)$, being a chiral primary with $h = \frac{3}{2}$, is a section of a line bundle $K \otimes \mathcal{L} \equiv \mathcal{L}^3$ where the line bundle $\mathcal{L}$ is a *chosen* spin-structure on Σ (cf. **BOX 1.10**).

⁷¹ In superstring theory $d = 10$, or, if we work in the light-cone gauge, where only the transverse fermions ψ^i survive, $d = 8$.

Let $\epsilon(z)$ be a holomorphic (anticommuting) spinor depending holomorphically on z . More precisely, $\epsilon(z)$ should be a holomorphic section of $\mathcal{L}^{-1}$, that is, a 2d *conformal Killing spinor* (cf. Eq. (1.133)). Then the chiral currents

$$J(z) = \epsilon(z) T_F(z), \quad \tilde{J}(\bar{z}) = \bar{\epsilon}(\bar{z}) \tilde{T}_F(\bar{z}) \quad (2.396)$$

are *globally defined* conserved currents of weights $(1, 0)$ and $(0, 1)$, respectively, i.e. they are (anti-)holomorphic 1-forms which may be integrated along closed contours $C \subset \Sigma$ to define conserved *supercharges* $Q^\epsilon(C)$, $\tilde{Q}^{\bar{\epsilon}}(C)$ acting on the state/operator encircled by C . The supercharge $Q^\epsilon(C)$ implements the SUSY transformation of states and operators of Grassman parameter $\epsilon(z)$.

Example: the Free $(1, 1)$ SCFT

A basic example is the free $(1, 1)$ SCFT corresponding to the matter sector of the world-sheet theory of the critical superstring in the conformal gauge (we set $\alpha' = 2$):

$$S = \frac{1}{4\pi} \int d^2z \left(\partial X^\mu \bar{\partial} X_\mu + \psi^\mu \bar{\partial} \psi_\mu + \tilde{\psi}^\mu \partial \tilde{\psi}_\mu \right), \quad (2.397)$$

where the free Majorana–Weyl fermions ψ_μ (resp. $\tilde{\psi}_\mu$) are chiral conformal primaries of weights $(h, \tilde{h}) = (1/2, 0)$ (resp. $(0, 1/2)$), that is, ψ^μ is a section of a spin bundle $\mathcal{L}$ (resp. $\tilde{\psi}^\mu$ is a section of $\tilde{\mathcal{L}}$) and $\mu = 0, \dots, d-1$.

The OPE of the free fields in Eq. (2.397) are

$$X^\mu(z, \bar{z}) X^\nu(0, 0) \sim -\eta^{\mu\nu} \log |z|^2, \quad (2.398)$$

$$\psi^\mu(z) \psi^\nu(0) \sim \frac{\eta^{\mu\nu}}{z}, \quad \tilde{\psi}^\mu(\bar{z}) \tilde{\psi}^\nu(0) \sim \frac{\eta^{\mu\nu}}{\bar{z}}. \quad (2.399)$$

The supercurrents are

$$T_F(z) = i \psi^\mu(z) \partial X_\mu(z), \quad \tilde{T}_F(\bar{z}) = i \tilde{\psi}^\mu(\bar{z}) \bar{\partial} X_\mu(\bar{z}), \quad (2.400)$$

while the left-moving energy–momentum tensor⁷³ is the sum of the tensors for the free bosons and the free fermions

$$T_B(z) = -\frac{1}{2} \partial X^\mu \partial X_\mu - \frac{1}{2} \psi^\mu \partial \psi_\mu, \quad (2.401)$$

The two chiral currents T_B and T_F form a chiral algebra (called the $\mathcal{N} = 1$ *superconformal algebra*) which is best written in terms of OPEs

⁷² The spin-structures for left- and right-movers need not to be the same, i.e. we may have $\mathcal{L} \not\cong \mathcal{L}'$.

⁷³ In the superconformal context the ordinary energy–momentum tensor is written as $T_B(z)$ to emphasize that it is the bosonic component of the energy–momentum *superfield* $\mathbf{T}(z, \theta) = T_F(z) + \theta T_B(z)$ which also has a fermionic component, namely the supercurrent $T_F(z)$.

$$T_B(z) T_B(0) \sim \frac{3d}{4z^4} + \frac{2}{z^2} T_B(0) + \frac{1}{z} \partial T_B(0) \quad (2.402)$$

$$T_B(z) T_F(0) \sim \frac{3}{2z^2} T_F(0) + \frac{1}{z} \partial T_F(0) \quad (2.403)$$

$$T_F(z) T_F(0) \sim \frac{d}{z^3} + \frac{2}{z} T_B(0). \quad (2.404)$$

Equation (2.402) is just the Virasoro algebra with central charge $3d/2$ where d is the number of dimensions of the target space: each boson contributes $+1$ to c and each (real) fermion contributes $+1/2$. Equation (2.403) says that T_F is a Virasoro primary operator with weights $(3/2, 0)$ (as expected). Equation (2.404) says that the anticommutator of two superconformal transformations is an infinitesimal conformal motion up to a central extension. The central term in Eq. (2.404) is related to the central term c in (2.402) by the superJacobi identities (equivalently: by associativity of the OPE algebra).

The Superconformal Algebra

From the arguments in the previous paragraph, we see that the superconformal algebra of *any* $\mathcal{N} = 1$ SCFT should have the same structure as in the example (2.402)–(2.404) since the first two equations just reflect the fact that we have a CFT with a spin- $\frac{3}{2}$ chiral current and the third one follows from the first two by associativity of the OPE algebra. The $\mathcal{N} = 1$ superconformal algebra (written in terms of OPEs) then reads

$$T_B(z) T_B(0) \sim \frac{3\hat{c}}{4z^4} + \frac{2}{z^2} T_B(0) + \frac{1}{z} \partial T_B(0) \quad (2.405)$$

$$T_B(z) T_F(0) \sim \frac{3}{2z^2} T_F(0) + \frac{1}{z} \partial T_F(0) \quad (2.406)$$

$$T_F(z) T_F(0) \sim \frac{\hat{c}}{z^3} + \frac{2}{z} T_B(0), \quad (2.407)$$

where

$$\hat{c} \equiv \frac{2c}{3} \quad (2.408)$$

with c the Virasoro central charge. $\hat{c}$ is normalized to be 1 for each free $\mathcal{N} = 1$ supermultiplet (2.410) consisting of a real scalar and a Majorana fermion.

2.8.1 Primary Superfields

In a SCFT the *primary superfields* are superconformal tensor fields $\Phi(z, \theta)$ on *super-Riemann surfaces*.⁷⁴ In a local coordinate patch of holomorphic coordinates (z, θ) ,

⁷⁴ A *super-Riemann surface* is a complex supermanifold of dimension $1|1$ with a maximally non-integrable $0|1$ sub-bundle $\mathcal{D}$ of the tangent bundle. The last statement means that locally there is an

where z is a “usual” commuting complex coordinate and θ is an anticommuting (Grassman) coordinate, a *superconformal superfield*⁷⁵ of weight h has the form

$$\Phi(z, \theta) = \phi_0(z) + \theta \phi_1(z), \quad (2.409)$$

where ϕ_0, ϕ_1 are ordinary (Virasoro) conformal primaries of weights h and $h + \frac{1}{2}$, respectively, with *opposite statistics*. For instance, for the free theory in Eq. (2.397), we have the chiral superconformal field of weight $h = \frac{1}{2}$ and Fermi statistics⁷⁶

$$DX^\mu \equiv -i \psi^\mu(z) + \theta \partial X^\mu(z), \quad (2.410)$$

where D is the *superderivative*

$$D = \frac{\partial}{\partial \theta} + \theta \frac{\partial}{\partial z}, \quad D^2 = \partial_z. \quad (2.411)$$

For a general conformal superfield $\Phi(z, \theta)$ of weight h we have

$$T_B(z) \phi_0(0) \sim \frac{h}{z^2} \phi_0(0) + \frac{1}{z} \partial \phi_0 \quad (2.412)$$

$$T_B(z) \phi_1(0) \sim \frac{h + 1/2}{z^2} \phi_1(0) + \frac{1}{z} \partial \phi_1 \quad (2.413)$$

$$T_F(z) \phi_0(0) \sim \frac{1}{z} \phi_1(0) \quad (2.414)$$

$$T_F(z) \phi_1(0) \sim \frac{2h}{z^2} \phi_0(0) + \frac{1}{z} \partial \phi_0, \quad (2.415)$$

Equations (2.412), (2.413) just say that the component fields $\phi_0(z), \phi_1(z)$ are Virasoro primaries of weights $h, h + \frac{1}{2}$, respectively. Equation (2.414) says that the SUSY transformation of the first (lowest) component of the supermultiplet, ϕ_0 , is the second (highest) component ϕ_1 . Finally (2.415) says that the anticommutator of two superconformal symmetries is a conformal transformation; see Eq. (2.404). Equation (2.415) follows from the previous ones by associativity of the OPE algebra. Indeed taking into account the $z \leftrightarrow w$ symmetry, the OPE (2.407), and the order of poles as $z \rightarrow w, z \rightarrow 0$, and $w \rightarrow 0$, we get

$$T_F(z) T_F(w) \phi_0(0) \sim \frac{\hat{c} \phi_0(0)}{(z-w)^3} + \frac{2h \phi_0(0)}{(z-w)zw} + \frac{\partial \phi(0)}{(z-w)z} + \frac{\partial \phi(0)}{(z-w)w} \quad (2.416)$$

odd first-order differential operator D (called the *superderivative*; cf. Eq. (2.411)) such that D^2 is a nowhere vanishing *even* first-order differential operator which we may locally identify with ∂_z .

⁷⁵ We write only the holomorphic part, the anti-holomorphic sector being similar and independent of the holomorphic one.

⁷⁶ The statistics of a superfield is, by convention, the statistics of its first component.

and hence

$$T_F(z) \phi_1(0) = \lim_{w \rightarrow 0} \left(w T_F(z) T_F(w) \phi_0(0) \right) \sim \frac{2h \phi_0(0)}{z^2} + \frac{\partial \phi_0(0)}{z}. \quad (2.417)$$

2.8.2 Ramond and Neveu–Schwarz Sectors

We study a $\mathcal{N} = 1$ SCFT quantized on a circle S^1 of length 2π (the Euclidean world-sheet is then the infinite cylinder $S^1 \times \mathbb{R}$). We consider first the deflist in terms of the cylindrical coordinate

$$w = \sigma^1 + i\sigma^2, \quad \text{with } w \simeq w + 2\pi. \quad (2.418)$$

All bosonic local operators are periodic in w ,

$$\mathcal{O}_B(w + 2\pi) = \mathcal{O}_B(w). \quad (2.419)$$

Since bilinears in Fermi operators are bosonic, there are two periodicity conditions on the fermions which are consistent with (2.419), namely *periodic* or *anti-periodic*. They correspond to the two inequivalent spin-structures on the cylinder.⁷⁷ In particular, for the supercurrent we have

$$T_F(w + 2\pi) = \exp(2\pi i \nu) T_F(w) \quad \text{with } \nu = 0, \frac{1}{2}. \quad (2.420)$$

The left-moving Hilbert space on S^1 then consists of two sectors:

- the one with $\nu = 0$ called the *Ramond* sector (R);
- the one with $\nu = \frac{1}{2}$ called the *Neveu–Schwarz* sector (NS).

(1, 1) SCFT on a Cylinder For a (1, 1) SCFT on a cylinder the possible periodicity conditions are

$$\begin{aligned} T_F(w + 2\pi) &= \exp(2\pi i \nu) T_F(w) \\ \tilde{T}_F(\bar{w} + 2\pi) &= \exp(-2\pi i \tilde{\nu}) \tilde{T}_F(\bar{w}) \quad \text{with } \nu, \tilde{\nu} = 0, \frac{1}{2}. \end{aligned} \quad (2.421)$$

Therefore there are *four* different ways to quantize a (1,1) SCFT on a circle: each one leads to a different sector of the Hilbert space $\mathbf{H}_{S^1}$. We denote these 4 sectors of the SCFT as $(\nu, \tilde{\nu})$ or as NS-NS, NS-R, R-NS, and R-R, respectively. In the NS sector the two component fields ϕ_0, ϕ_1 have opposite periodicity on the cylinder.

⁷⁷ A spin-structure on a flat Σ reduces to a group homomorphism $H_1(\Sigma, \mathbb{Z}) \rightarrow \{\pm 1\}$. For the cylinder $H_1(\mathbf{C}y, \mathbb{Z}) \cong \mathbb{Z}$, and we have two spin-structures: $n \mapsto 1$ (periodic) and $n \mapsto (-1)^n$ (anti-periodic).

For instance, for the free model (2.397) the boundary conditions consistent with superconformal invariance are

$$\psi^\mu(w + 2\pi) = \exp(2\pi i \nu) \psi^\mu(w) \quad (2.422)$$

$$\tilde{\psi}^\mu(\bar{w} + 2\pi) = \exp(-2\pi i \tilde{\nu}) \tilde{\psi}^\mu(\bar{w}), \quad (2.423)$$

where $\nu, \tilde{\nu}$ take the values 0 or $\frac{1}{2}$, and are the same for all μ since they are fixed by the periodicity of the supercurrents, Eq. (2.421). In the superstring case the independence of $\nu, \tilde{\nu}$ from the index μ also follows from the spacetime Poincaré symmetry.

(1, 1) SCFT on a Strip

Next we take Σ to be the strip $[0, \pi] \times \mathbb{R}$ with coordinate $w = \sigma^1 + i\sigma^2, 0 \leq \text{Re } w \leq \pi$. Consistency of the theory as a CFT requires the boundary conditions on the energy–momentum tensor (see Eq. (2.309))

$$T_B(w) = \tilde{T}_B(\bar{w}) \text{ for } \text{Re } w = 0, \pi. \quad (2.424)$$

Consistency of the SCFT OPE (2.407) then requires the following boundary conditions on the supercurrent

$$\begin{aligned} T_F(0, \sigma^2) &= \exp(2\pi i \nu) \tilde{T}_F(0, \sigma^2), \\ T_F(\pi, \sigma^2) &= \exp(2\pi i \nu') \tilde{T}_F(\pi, \sigma^2), \end{aligned} \quad (2.425)$$

with $\nu, \nu' \in \{0, \frac{1}{2}\}$. By the redefinition $\tilde{T}_F \rightarrow e^{-2\pi i \nu'} \tilde{T}_F$ we set $\nu' = 0$. We are left with two distinct sectors, R and NS, for $\nu = 0$ or, respectively, $\nu = \frac{1}{2}$. We combine T_F and $\tilde{T}_F$ into a single supercurrent in the doubled range $0 \leq \sigma^1 \leq 2\pi$ by setting

$$T_F(\sigma^1, \sigma^2) = \begin{cases} T_F(\sigma^1, \sigma^2) & \text{for } 0 \leq \sigma^1 \leq \pi \\ \tilde{T}_F(2\pi - \sigma^1, \sigma^2) & \text{for } \pi \leq \sigma^1 \leq 2\pi. \end{cases} \quad (2.426)$$

The boundary condition at $\sigma^1 = \pi$ becomes the continuity of $T_F(z)$ across the real axis. The anti-holomorphicity of $\tilde{T}_F$ guarantees the holomorphicity of $T_F(w)$ in its full domain (by the Schwarz reflection principle [48] a.k.a. the *doubling trick*; cf. Sect. 2.6). Finally, the boundary condition at $\sigma^1 = 0$ becomes a periodicity condition on the extended supercurrent $T_F(w)$

$$T_F(w + 2\pi) = \exp(2\pi i \nu) T_F(w), \quad (2.427)$$

giving *one single copy* of the chiral superconformal algebra (2.405)–(2.407).

In a (1, 1) SCFT on a cylinder we have two copies (left and right) of the superconformal algebra. For the rest of this section we focus on one copy of the algebra.

Mode Expansions

In the cylindrical coordinate w we have the Fourier expansions⁷⁸

$$T_F(w) = i^{-3/2} \sum_{r \in \mathbb{Z} + \nu} e^{irw} G_r, \quad \tilde{T}_F(\bar{w}) = i^{3/2} \sum_{r \in \mathbb{Z} + \tilde{\nu}} e^{-ir\bar{w}} \tilde{G}_r, \quad (2.428)$$

where ν (resp. $\tilde{\nu}$) is equal to 0 for the R-sector and $\frac{1}{2}$ for the NS one.

Now we go to the more convenient radial quantization, where the mode expansion takes the form of a Laurent series. We set $z = \exp(-iw)$. Taking into account the conformal weight of the supercurrents,

$$\begin{aligned} T_F(z) (dz)^{3/2} &= T_F(w) (dw)^{3/2} \Rightarrow \\ \Rightarrow T_F(z) &= (\partial w / \partial z)^{3/2} T_F(w) = i^{3/2} z^{-3/2} T_F(w), \end{aligned} \quad (2.429)$$

the Laurent expansions for $T_B(z)$ and $T_F(z)$ become

$$T_F(z) = \sum_{r \in \mathbb{Z} + \nu} \frac{G_r}{z^{r+3/2}}, \quad \tilde{T}_F(\bar{z}) = \sum_{r \in \mathbb{Z} + \tilde{\nu}} \frac{\tilde{G}_r}{\bar{z}^{r+3/2}}, \quad (2.430)$$

$$T_B(z) = \sum_{m \in \mathbb{Z}} \frac{L_m}{z^{m+2}}, \quad \tilde{T}_B(\bar{z}) = \sum_{m \in \mathbb{Z}} \frac{\tilde{L}_m}{\bar{z}^{m+2}}. \quad (2.431)$$

We stress that the indices r, s are *integral* in the Ramond sector and *half-integral* in the Neveu–Schwarz one.

Inserting the expansions (2.430), (2.431) in Eqs. (2.405)–(2.407) yields the (1, 1) superconformal algebra written in terms of modes

$$[L_m, L_n] = (m - n) L_{m+n} + \frac{c}{12} (m^3 - m) \delta_{m, -n} \quad (2.432)$$

$$\{G_r, G_s\} = 2 L_{r+s} + \frac{c}{12} (4r^2 - 1) \delta_{r, -s} \quad (2.433)$$

$$[L_m, G_r] = \frac{m - 2r}{2} G_{m+r}. \quad (2.434)$$

This infinite-dimensional Lie superalgebra is known as *the Ramond algebra* for r, s integral and as the *Neveu–Schwarz algebra* for r, s half-integral. The anti-holomorphic sector gives a second copy of these algebras. Notice that the R and NS algebras are *not* isomorphic.

Note 2.8 $L_0, L_{\pm 1}, G_{\pm 1/2}$ form a finite-dimensional sub-superalgebra of the NS algebra, which corresponds to the superconformal symmetry $OSp(1|2)$ of the sphere; cf. **BOX 1.11**. Likewise, G_0 and $L_0 - c/24$ form a finite-dimensional sub-superalgebra of the R algebra, namely the rigid supersymmetry algebra of the flat cylinder.

⁷⁸ The conventional overall phases are introduced for later convenience.

Superfields For a (conformal) superfield we have the mode expansion

$$\Phi(z, \theta) = \sum_n \frac{\phi_{0,n}}{z^{n+h}} + \theta \sum_n \frac{\phi_{1,n}}{z^{n+h+1/2}}, \quad (2.435)$$

where for the spinorial component n takes values in $\mathbb{Z}$ or in $\mathbb{Z} + \frac{1}{2}$ for the R and resp. NS sector. We stress that on the plane (i.e. in radial quantization) the NS sector spinor fields are *univalued* (i.e. periodic as $z \rightarrow e^{2\pi i} z$) while the R-sector ones are *doubly valued* (anti-periodic in the angular variable).

From Eqs. (2.412)–(2.415) one has⁷⁹

$$[L_m, \phi_0(z)] = z^{m+1} \partial \phi_0 + h(m+1) z^m \phi_0(z), \quad (2.436)$$

$$[\epsilon G_m, \phi_0(z)] = \epsilon z^{m+1/2} \phi_1(z), \quad (2.437)$$

$$[L_m, \phi_1(z)] = z^{m+1} \partial \phi_1(z) + (h+1/2)(m+1) z^m \phi_1(z), \quad (2.438)$$

$$[\epsilon G_m, \phi_1(z)] = \epsilon z^{m+1/2} \partial \phi_0 + 2\epsilon(m+1/2) h z^{m-1/2} \phi_0(z), \quad (2.439)$$

or, in terms of the modes (2.435),

$$[L_m, \phi_{0,n}] = ((h-1)m - n) \phi_{0,m+n}, \quad (2.440)$$

$$[L_m, \phi_{1,n}] = ((h - \frac{1}{2})m - n) \phi_{1,m+n}, \quad (2.441)$$

$$[\epsilon G_m, \phi_{0,n}] = \epsilon \phi_{1,m+n}, \quad (2.442)$$

$$[\epsilon G_m, \phi_{1,n}] = \epsilon ((2h-1)m - n) \phi_{0,m+n}. \quad (2.443)$$

Note 2.9 The above superfields $\Phi(z, \theta)$ are *conformal* superfields not ordinary ones: they carry a (field) representation of the full *superconformal* SUSY algebra not just of the Poincaré one. In particular, they have a definite dimension h given by the Virasoro weight of their lowest component $\phi_0(z)$.

2.8.3 SCFT State-Operator Correspondence

NS Primaries

The first components $\phi_0(z)$ of the primary superfields form a special class of operators of the SCFT. Under the CFT operator-state correspondence, the lowest components of conformal superfields, $\phi_0(z)$, are mapped to *NS superconformal highest weight states* (or *NS primaries*, for short). These states belong to the NS sector and satisfy

$$G_r |\phi_0\rangle = 0 \quad r \geq \frac{1}{2} \implies L_n |\phi_0\rangle = 0 \quad n \geq 1. \quad (2.444)$$

⁷⁹ Here ϵ is an anticommuting parameter introduced to convert anticommutators into commutators.

Note that the superconformal algebra implies that all the conditions (2.444) are satisfied iff the following two relations hold

$$G_{1/2}|\phi_0\rangle = G_{3/2}|\phi_0\rangle = 0. \quad (2.445)$$

These equations follow directly from the definition of the superconformal superfields. (See discussion below.)

Spin Fields: World-Sheet SUSY

The ordinary (Poincaré) SUSY algebra in the NS sector is

$$G_{-1/2}^2 = L_{-1}, \quad (2.446)$$

with L_{-1} the generator of translations in the plane, $z \rightarrow z + \text{const}$. In the R-sector the Poincaré SUSY algebra reads

$$G_0^2 = L_0 - \hat{c}/16, \quad (2.447)$$

where the RHS is the translation generator on the cylinder, i.e. the left-moving Hamiltonian H_L of the theory on $S^1 \times \mathbb{R}$. The operator $G_0^2 \equiv G_0^\dagger G_0$ is non-negative in a unitary SCFT, so in the R-sector we have the *2d BPS bound*

$$h \geq \hat{c}/16 > 0. \quad (2.448)$$

SUSY is unbroken iff there are Ramond ground states $|\Omega\rangle$ which saturate the bound

$$G_0|\Omega\rangle = 0 \iff h = \hat{c}/16. \quad (2.449)$$

In a unitary SCFT the conformal vacuum $|0\rangle$ (the state corresponding to the identity) is the lowest energy state with $h = 0$. Since all states in the R-sector obey the bound (2.448), the vacuum $|0\rangle$ belongs to the NS sector and is invariant under the global superconformal super-group $OSp(2|1)$ generated by L_0 , $L_{\pm 1}$, and $G_{\pm 1/2}$.

The conformal superfields $\Phi_j(z, \theta) = \phi_{j,0}(z) + \theta\phi_{j,1}(z)$ of weight h_j create all the NS highest weight states

$$|h_j\rangle = \Phi_j(0, 0)|0\rangle \equiv \phi_{j,0}(0)|0\rangle, \quad (2.450)$$

which are annihilated by all superconformal lowering operators

$$G_r|h_j\rangle = L_n|h_j\rangle = 0 \quad \text{for } r, n > 0, \quad (2.451)$$

because of Eqs. (2.440)–(2.443).

However, the superfields $\Phi_j(z, \theta)$ **do not exhaust** the set of local operators of the theory: they cannot create states in the Ramond sector since the superfields do not modify the boundary conditions on the fermions. The states of the Ramond

sector are created by *other* conformal fields, called *spin fields*, which *should* exist by the CFT state/operator correspondence. The OPE of a spin field with the fermionic part of a NS superfield must be *non-local* (i.e. double-valued) in order to generate the correct mode expansions for the Ramond sector. Spin fields flip the boundary condition for the fermions between periodic and anti-periodic. Thus a spin field may be represented as the *endpoint of a branch cut* in the fermionic field just as the order/disorder operators in the Ising model [67, 68] which are the historically original examples of spin fields.⁸⁰

Since G_0 and L_0 commute, the spin fields come in pairs $S^\pm(z)$ such that

$$|h^\pm\rangle = S^\pm(0)|0\rangle, \quad |h^-\rangle = G_0|h^+\rangle, \quad (h - \hat{c}/16)|h^+\rangle = G_0|h^-\rangle, \quad (2.452)$$

that is, we have the OPE

$$T_F(z) S^\pm(w) \sim \frac{1}{(z-w)^{3/2}} a_\pm S^\mp(w), \quad \text{with } \begin{cases} a_+ = 1 \\ a_- = h - \hat{c}/16. \end{cases} \quad (2.453)$$

If some R-sector state has $h^+ = \hat{c}/16$, global SUSY is unbroken in the R-sector, and the ground states need not to be paired up—i.e. the corresponding ket $|h^-\rangle$ may be the zero vector. The 2d *Witten index* $\text{Tr}(-1)^F$ of the SCFT counts the net number of such unpaired Ramond ground states [69].

The general state in the R-sector is created by acting with a NS superfield on a R ground state (generated by a spin field of lowest dimension).

Mutual Locality: The GSO Projection

Two local operators $O_1(z, \bar{z})$ and $O_2(w, \bar{w})$ are said to be *mutually local* if their OPE

$$O_1(z, \bar{z}) O_2(w, \bar{w}) \quad (2.454)$$

is univalued in the plane, i.e. its analytic continuation along a curve $z = e^{2\pi i t} r + w$ ($t \in \mathbb{R}$) which encircles the point w is periodic in t of period 2π . Two operators O_α and O_β are mutually local iff for all operator O_γ entering in its OPE we have

$$(h_\alpha - \tilde{h}_\alpha) + (h_\beta - \tilde{h}_\beta) - (h_\gamma - \tilde{h}_\gamma) \in \mathbb{Z}. \quad (2.455)$$

In particular, the mutual locality of $O(z, \bar{z})$ with respect to its own Hermitian conjugate $O(z, \bar{z})^\dagger$ requires its conformal spin $h - \tilde{h}$ to be either integral or half-integral.

A CFT is said to be *local* if all its local operators are pair-wise mutually local. In a local CFT all correlation functions

$$\langle O_1(z_1, \bar{z}_1) \cdots O_s(z_s, \bar{z}_s) \rangle_\Sigma \quad (2.456)$$

⁸⁰ Since the Ising model at criticality is equivalent to the CFT given by a free massless Majorana fermion λ , the order/disorder operators are just the spin fields, which creates branch cut in λ .

are well-defined univalued functions (or rather distributions) of the positions $z_j \in \Sigma$ of the inserted local operators. A *necessary* condition for a CFT to be local is $h - \tilde{h} \in \frac{1}{2}\mathbb{Z}$ for all local operators.

In a unitary⁸¹ $\mathcal{N} = 1$ SCFT with $h - \tilde{h} \in \frac{1}{2}\mathbb{Z}$, the full set of all NS and R operators **cannot** correspond to a *local* SCFT since (by definition) the fermionic fields are doubly valued around the spin fields. Locality on the world-sheet is crucial in order to have a well-defined (perturbative) superstring theory: the correlations function must be integrated over the positions⁸² of the inserted operators (to enforce invariance under 2d diffeomorphisms) and the integral of a multivalued function has no meaning.

There are two ways to extract from the non-local $\text{NS} \oplus \text{R}$ operator algebra a local subalgebra. The first way is to restrict to the NS sector. (The restriction to the R-sector is inconsistent: the OPE does not close because the OPE of two spin fields is a NS operator). The second way, the *GSO projection*,⁸³ keeps both sectors: after eliminating half of each, one gets a local operator algebra, hence a local SCFT.

GSO Projection The fermion parity operator

$$(-1)^F \equiv \Gamma \quad (2.457)$$

which counts fermions mod 2, anticommutes with all the fermionic components of the superfields and commutes with their bosonic parts. It is well-known that the ± 1 -eigenspaces of Γ

$$\mathbf{H} = \mathbf{H}_+ \oplus \mathbf{H}_- \quad (2.458)$$

are superselected sectors of the Hilbert space $\mathbf{H}$.

In the context of superstring theory, Γ is called the *chirality operator* because it measures the chirality of physical states in the spacetime sense.⁸⁴

Spin fields of opposite chirality are mutually **non**-local since their OPE, being a Γ -odd NS field, should contain an odd number of fermions.

In a unitary theory with $h - \tilde{h} \in \frac{1}{2}$, spin fields of the *same* chirality are mutually local between themselves and also to the bosonic fields, including the Fermi bilinears. Indeed these spin fields have all the same conformal spin mod 1 and their OPE closes on Γ -even (i.e. bosonic) NS operators which have integral conformal spin, so that Eq. (2.455) is satisfied. Therefore the projection into the $\Gamma = 1$ sector $\mathbf{H}_+$ yields a local SCFT. The fields which survive the projection are the bosonic parts of the NS superfields, including fermion bilinears, and the spin fields of positive chirality.

⁸¹ Unitarity requires that NS operators with $h - \tilde{h}$ integral (half-integral) are bosons (resp. fermions).

⁸² To be precise, the integral is over the moduli $\mathcal{M}_{g,n}$ of punctured surfaces. The correlations are univalued in $\mathcal{M}_{g,n}$ only if the inserted operators are mutually local.

⁸³ The *GSO projection* is named after Gliozzi, Scherk, and Olive [70].

⁸⁴ For a detailed discussion of this topic, see Sect. 3.1.

2.8.4 Example: The Free SCFT

We quantize the free SCFT (2.397), corresponding to the matter part of the superstring world-sheet theory, on a circle of length 2π (this is the closed string sector).

Mode Expansion on a Circle We expand in Fourier modes on the cylinder

$$\psi^\mu(w) = i^{-1/2} \sum_{r \in \mathbb{Z} + \nu} \psi_r^\mu \exp(irw), \quad \tilde{\psi}^\mu(\bar{w}) = i^{1/2} \sum_{r \in \mathbb{Z} + \bar{\nu}} \tilde{\psi}_r^\mu \exp(-ir\bar{w}) \quad (2.459)$$

the overall phase being chosen to agree with standard conventions. The sum runs over the integers for the R-sector and over the integers plus $+\frac{1}{2}$ for the NS sector.

Now we go to radial quantization, where the mode expansion takes the form of a Laurent series (instead of a Fourier one as in cylindrical coordinates). We set $z = \exp(-iw)$. Taking into account the weight of the fields we get

$$\psi^\mu(z) = \sum_{r \in \mathbb{Z} + \nu} \frac{\psi_r^\mu}{z^{r+1/2}}, \quad \tilde{\psi}^\mu(\bar{z}) = \sum_{r \in \mathbb{Z} + \bar{\nu}} \frac{\tilde{\psi}_r^\mu}{\bar{z}^{r+1/2}}. \quad (2.460)$$

Notice that in the NS sector the branch cut $z^{-1/2}$ eliminates the original anti-periodicity, while in the R-sector the fermionic fields get a branch cut. Let us recall the corresponding bosonic expansions

$$\partial X^\mu = -i \sum_{m=-\infty}^{+\infty} \frac{\alpha_m^\mu}{z^{m+1}}, \quad \bar{\partial} X^\mu = -i \sum_{m=-\infty}^{+\infty} \frac{\tilde{\alpha}_m^\mu}{\bar{z}^{m+1}} \quad (2.461)$$

where $\alpha_0^\mu = \tilde{\alpha}_0^\mu = p^\mu$ in the closed string and $\alpha_0^\mu = 2p^\mu$ in the open string (here we set $\alpha' = 2$ to simplify the expressions).

From the OPE and the mode expansions we get the canonical (anti) commutators

$$\{\psi_r^\mu, \psi_s^\nu\} = \{\tilde{\psi}_r^\mu, \tilde{\psi}_s^\nu\} = \eta^{\mu\nu} \delta_{r,-s} \quad (2.462)$$

$$[\alpha_m^\mu, \alpha_n^\nu] = [\tilde{\alpha}_m^\mu, \tilde{\alpha}_n^\nu] = m \eta^{\mu\nu} \delta_{m,-n}. \quad (2.463)$$

For the free SCFT (2.397) we may write the superconformal generators L_m, G_r in terms of the modes of the fields X^μ, ψ^μ as

$$L_m = \frac{1}{2} \sum_{n \in \mathbb{Z}} \alpha_{m-n}^\mu \alpha_{\mu n} + \frac{1}{4} \sum_{r \in \mathbb{Z} + \nu} (2r - m) \psi_{m-r}^\mu \psi_{\mu r} + a^M \delta_{m,0} \quad (2.464)$$

$$G_r = \sum_{n \in \mathbb{Z}} \alpha_n^\mu \psi_{\mu r-n}. \quad (2.465)$$

Here $\circ \circ$ stands for the *standard*⁸⁵ creation–annihilation normal order (not to be confused with the *conformal* one written as $: \cdots :$).

The “normal ordering” constant a^M may be computed in various ways, e.g. as the zero-point energy ($\equiv$ Casimir energy on a circle of length 2π). From **BOX 1.2** we see each periodic boson contributes $-1/24$ each periodic fermion $+1/24$ and an anti-periodic fermion $-1/48$. Including the shift by $+c/24 = d/16$ we get

$$a^M \Big|_R = \frac{d}{16}, \quad a^M \Big|_{NS} = 0. \quad (2.466)$$

Mode Expansion on an Interval

Quantizing (2.397) on the interval $[0, \pi]$, the condition that the boundary terms in the equation of motion vanish allows for the possibilities

$$\psi^\mu(0, \sigma^2) = \exp(2\pi i \nu) \tilde{\psi}^\mu(0, \sigma^2) \quad (2.467)$$

$$\psi^\mu(\pi, \sigma^2) = \exp(2\pi i \nu') \tilde{\psi}^\mu(\pi, \sigma^2). \quad (2.468)$$

By a redefinition $\tilde{\psi}^\mu \rightarrow e^{-2\pi i \nu'} \tilde{\psi}^\mu$, we may set $\nu' = 0$. We are left with just two sectors, R and NS, for $\nu = 0$ and, respectively, $\nu = \frac{1}{2}$. To write the mode expansion, it is convenient to combine ψ^μ and $\tilde{\psi}^\mu$ into a single field in the doubled range $0 \leq \sigma^1 \leq 2\pi$ by setting

$$\psi^\mu(\sigma^1, \sigma^2) = \tilde{\psi}^\mu(2\pi - \sigma^1, \sigma^2) \quad \text{for } \pi \leq \sigma^1 \leq 2\pi. \quad (2.469)$$

The boundary condition at $\sigma^1 = \pi$ is then the continuity of ψ^μ ; the anti-holomorphicity of $\tilde{\psi}^\mu$ guarantees the holomorphicity of ψ^μ in its full domain (again by the Schwarz reflection principle [48]). Finally, the boundary condition at $\sigma^1 = 0$ becomes a periodicity condition on the extended field $\psi^\mu(z)$, leaving us with *one set* of R or NS oscillators ψ_r^μ, α_n^μ and the corresponding algebras (2.462), (2.463).

In the rest of the section we focus on a single chiral sector: there are two copies (left and right) in the closed superstring and one copy in the open superstring.

NS and R Spectra

We consider the spectrum generated by a single set of modes (corresponding to the open superstring or one side of the closed one). The bosonic modes α_m^μ act as in the bosonic string. In the NS sector ψ^μ has no $r = 0$ mode, and the ground state is the unique state annihilated by all $r > 0$ modes

$$\psi_r^\mu |0\rangle_{NS} = 0, \quad r > 0, \quad (2.470)$$

while the $r < 0$ modes are creation operators. Being anticommuting, each mode ψ_{-r}^μ may be excited only once.

⁸⁵ That is, the positive modes are written to the right of the negative modes, independently of the value of the conformal spin of the field.

Consider now the R-sector. From Eq.(2.462) we see that the zero-modes ψ_0^μ satisfy the Dirac matrix algebra

$$\{\Gamma^\mu, \Gamma^\nu\} = 2\eta^{\mu\nu} \quad \text{where} \quad \Gamma^\mu \stackrel{\text{def}}{=} \sqrt{2}\psi_0^\mu. \quad (2.471)$$

This entails that the R vacuum is degenerate since

$$\{\psi_r^\mu, \psi_0^\nu\} = 0 \quad \text{for } r > 0, \quad (2.472)$$

so that the operators ψ_0^μ map R ground states into R ground states. Hence the R-sector vacua form a Majorana spinor.⁸⁶ with respect to the spacetime Lorentz symmetry.

In the R-sector not only the vacua but all states have half-integral spacetime spin because the raising operators ψ_{-r}^μ are vectors which change the spacetime spin by integers. In $d = 10$ (the critical dimension) the Dirac spinor representation has dimension **32**, and it is reducible into two Weyl representations

$$\mathbf{32} = \mathbf{16} \oplus \mathbf{16}' \quad (2.473)$$

distinguished by their eigenvalue under the spacetime chirality operator $\Gamma_{11} \equiv \Gamma$. The chirality quantum number Γ_{11} may be extended to the full superstring spectrum. The Dirac chiral matrix Γ_{11} is defined by the property of anticommuting with all matrices Γ^μ . Since the Dirac matrices are identified with the zero-modes of the fields $\psi^\mu(z)$, Eq.(2.471), the extension of Γ_{11} is a world-sheet operator which anticommutes with the full field $\psi^\mu(z)$. We call this operator

$$\exp(i\pi F), \quad (2.474)$$

where F is the (*world-sheet*) *fermion number* which for Majorana fermions is well-defined only mod 2. Since ψ^μ changes F by 1 mod 2, ψ^μ and $e^{i\pi F}$ anticommute. F may be written in terms of spacetime Lorentz generators for the ψ CFT

$$J^{\mu\nu} = -\frac{i}{2} \sum_{\mathbb{Z}+\nu} [\psi_r^\mu, \psi_{-r}^\nu]. \quad (2.475)$$

To see this, let us consider the Cartan generators

$$H_a = i^{-\delta_{a,0}} J^{2a, 2a+1} \quad a = 0, \dots, 4 \quad (2.476)$$

⁸⁶ In math parlance Eqs.(2.471), (2.472) say that the space of R vacua is a *Clifford module*. The Hermitian conjugate of a spin field with $h = \hat{c}/16$ is a spin field with the same weight, so the space of R vacua is a *real* module. Elements of real Clifford modules are called *Majorana spinors*.

(the extra i is needed to make H_0 Hermitian). Then

$$F = \sum_{a=0}^4 H_a. \quad (2.477)$$

Note that F is conserved in OPEs as a consequence of Lorentz invariance. We shall see in Chap. 3 that the superconformal ghosts also contribute to the Fermi number operator F , so that the chirality operator $\bar{\Gamma}$ relevant for the superstring will have an extra factor from the ghosts which redefines the chirality of the NS and R vacua.

The R vacua, being spinors, may be written in a basis of eigenstates of the Cartan generators H_a , i.e.

$$H_a |s_0, s_1, \dots, s_4\rangle_R = s_a |s_0, s_1, \dots, s_4\rangle_R \quad \text{where } s_a = \pm \frac{1}{2}. \quad (2.478)$$

Indeed such s_a are the weights of the spinorial representation of $\text{Spin}(9, 1)$; the spinor has chirality $+1$ (resp. -1) iff the number of $-\frac{1}{2}$'s in $(s_0, \dots, s_4)$ is even (resp. odd).

General Free SCFT

We may construct a free superconformal theory by combining an anticommuting b, c system with a commuting β, γ system. The NS superfields are

$$B(z) = \beta(z) + \theta b(z), \quad C(z) = c(z) + \theta \gamma(z), \quad (2.479)$$

with weights

$$\begin{aligned} h_b &= \lambda & h_c &= 1 - \lambda \\ h_\beta &= \lambda - \frac{1}{2} & h_\gamma &= \frac{3}{2} - \lambda. \end{aligned} \quad (2.480)$$

In particular, the superconformal ghosts are of this form. By definition, in this case γ (resp. c) has the quantum numbers of a conformal Killing spinor (resp. conformal Killing vector) and hence $h_\gamma = -\frac{1}{2}$, $h_c = -1$, that is, $\lambda = 2$. The action is (we write the left-movers only)

$$S = \frac{1}{2\pi} \int d^2z \left(b \bar{\partial} c + \beta \bar{\partial} \gamma \right) \quad (2.481)$$

and the superconformal currents are

$$T_B = (\partial b)c - \lambda \partial(bc) + (\partial\beta)\gamma - \left(\lambda - \frac{1}{2}\right)\partial(\beta\gamma) \quad (2.482)$$

$$T_F = -\frac{1}{2}(\partial\beta)c + \left(\lambda - \frac{1}{2}\right)\partial(\beta c) - 2b\gamma. \quad (2.483)$$

The Virasoro central charge is

$$c = \overbrace{1 - 3(2\lambda - 1)^2}^{bc} + \overbrace{-1 + 3(2(\lambda - 1/2) - 1)^2}^{\beta\gamma} = 9 - 12\lambda. \quad (2.484)$$

For the superconformal ghost system of the superstring, $\lambda = 2$, this yields $c = -15$ or $\hat{c} = -10$. The condition that the total central charge (of the matter system (2.397) and the ghost system (2.481)) gives the critical dimension for the superstring

$$d_{\text{crit.}} = 10. \quad (2.485)$$

2.9 $SO(2n)$ Current Algebra at Level 1 and Lattices

In this section we present the Frenkel–Kač–Segal construction of a simply laced untwisted affine Kač–Moody algebra at level 1 in terms of free bosons (and free fermions) focusing on the important example of $G = SO(2n)$. The construction for a general simply laced Lie algebra is essentially identical *mutatis mutandis*: we leave the cases A_n and E_r as an exercise for the reader.

2.9.1 The $SO(d-1, 1)$ World-Sheet Current Algebra

As the first motivation for our discussion (and its relevance for string theory), we consider the matter part of the superstring world-sheet SCFT, Eq. (2.397). We focus on the left-moving degrees of freedom; the right-moving story is similar.

From Eq. (2.471) we see that the zero-modes ψ_0^μ of the R-sector form a Clifford algebra in a d -dimensional space of signature $(d-1, 1)$. Since the zero-modes ψ_0^μ commute⁸⁷ with L_0 , they map R ground states into R ground states which then form a *Clifford module*, i.e. a spinor in $\mathbb{R}^{d-1,1}$. From the general discussion in Sect. 2.8.3, valid for all $(1, 1)$ SCFTs, we know that there exist spin fields $S_\alpha(z)$ which create these R-sector ground states from the NS vacuum $|0\rangle$ and transform as a spacetime spinor

$$|\alpha\rangle = S_\alpha(0)|0\rangle, \quad (2.486)$$

where $\alpha = 1, \dots, 2^{[d/2]}$ is a spinor index. $S_\alpha(z)$ has weight $h = \hat{c}/16 \equiv d/16$. Then a purely left-moving ground spin field of weights $(d/16, 0)$ has conformal spin in $\frac{1}{2}\mathbb{Z}$ (i.e. it is mutually local with respect to itself) iff d is a multiple of 8. $d = 8m$ are the dimensions where the Euclidean-signature Clifford algebras are isomorphic to *real* matrix algebras, in fact to the algebra of *real* $2^{4m} \times 2^{4m}$ matrices [71], i.e. they are precisely the dimensions where Euclidean-signature Majorana–Weyl spinors exist.

The *local* CFT obtained by projecting onto the even fermion number

$$\Gamma \equiv (-1)^F = 1 \quad (2.487)$$

⁸⁷ Indeed, from Eq. (2.436), $[L_0, \psi_r^\mu] = -r \psi_r^\mu$.

(*GSO projection*; cf. Sect. 2.8.3) is the world-sheet theory of the superstring. We will see in Sects. 3.1–3.3 that, after taking into account the superconformal Faddeev–Popov ghosts, it is natural to “effectively” assign to the *matter* NS vacuum $|0\rangle$ the fermion parity (chirality) $\Gamma_{11} = -1$. Hence the NS vacuum and all states created by acting on it with an *even* number of 2d fermionic fields get projected out. The lowest energy states in the NS sector which are not projected out are the states

$$\psi_{-1/2}^\mu |0\rangle \equiv DX^\mu(0, 0)|0\rangle. \quad (2.488)$$

We shall see that in critical dimension $d = 10$ these states are massless vectors.⁸⁸

The holomorphic currents

$$j^{\mu\nu}(z) \stackrel{\text{def}}{=} \psi^\mu \psi^\nu(z) \quad (2.489)$$

generate an $SO(d-1, 1)$ affine current algebra of level 1⁸⁹ which acts on the 2d fermionic degrees of freedom as the spacetime Lorentz algebra. Hence

$$j^{\mu\nu}(z) \psi^\lambda(w) \sim \frac{1}{z-w} (g^{\mu\lambda} \psi^\nu(w) - g^{\nu\lambda} \psi^\mu(w)), \quad (2.490)$$

$$j^{\mu\nu}(z) S_\alpha(w) \sim \frac{1}{z-w} (\gamma^{[\mu} \gamma^{\nu]}_\alpha)^\beta S_\beta(w), \quad (2.491)$$

where $(\gamma^\mu)_\alpha{}^\beta$ are Dirac matrices in d -dimensions. These relations imply the OPEs

$$\psi^\mu(z) S_\alpha(w) \sim (z-w)^{-1/2} \gamma_{\alpha\beta}^\mu S^\beta(w), \quad (2.492)$$

$$S_\alpha(z) S_\beta(w) \sim (z-w)^{-d/8} C_{\alpha\beta} + (z-w)^{1/2-d/8} (\gamma_{\mu})_{\alpha\beta} \psi^\mu(w) + (z-w)^{1-d/8} (\gamma_{\mu\nu})_{\alpha\beta} \psi^\mu \psi^\nu(w), \quad (2.493)$$

where $C_{\alpha\beta}$ is the charge conjugation matrix and $\gamma_{\mu\nu} \equiv \frac{1}{2}[\gamma_\mu, \gamma_\nu]$. Spinor indices are raised/lowered using the matrix $C_{\alpha\beta}$.

2.9.2 Bosonization of the $SO(2N)$ Current Algebra

To construct explicitly the superstring vertex operators,⁹⁰ it is convenient to *bosonize* the current algebra generated by the fermionic bilinears $\psi^\mu \psi^\nu(z)$.

⁸⁸ Compare with the discussion in Chap. 1 of the corresponding vector states in the bosonic string which are also massless for the same reason, i.e. because they have only $d-2$ physical polarizations.

⁸⁹ Indeed the central charge of the Fermi sector is $d/2$ which, for d even, is the rank of $SO(d-1, 1)$; compare with the discussion in Sect. 2.7.5.

⁹⁰ The explicit construction of the superstring vertices is given in Chap. 3 using the techniques developed in the present section.

We take $d = 2N$ to be *even*, and Wick rotate to the Euclidean Lorentz group $SO(2N)$. For $d = 2N$ the Lie algebra $\widetilde{\mathfrak{so}}(2N)$ is *simply laced*. $\psi^\mu(z), \tilde{\psi}^\mu(\bar{z})$ generate two copies of the affine Lie algebra $\widetilde{SO}(2N)$ at level 1. Indeed, the central charge of the free fermions ψ^μ is $c = 2N/2 = N$, equal to the rank of $SO(2N)$ (cf. Sect. 2.7.4). As already mentioned, all simply laced, level 1, affine Lie algebra admit an explicit construction in terms of free bosons due to Frenkel–Kač–Segal [64, 65]: we work out its details in the special case of $SO(2N)$.

It is convenient to rewrite the fermions in a complex basis

$$\lambda^{\pm e_j} \stackrel{\text{def}}{=} \sqrt{\frac{1}{2}} (\psi^{2j-1} \pm i \psi^{2j}), \quad j = 1, 2, \dots, N. \quad (2.494)$$

We identify the indices $+e_j$'s with the generators of the *lattice*⁹¹

$$\mathbb{Z}^N \equiv \left\{ \sum_{i=1}^N n_i e_i \mid n_i \in \mathbb{Z} \right\}, \quad (2.495)$$

endowed with the standard Euclidean inner product

$$e_i \cdot e_j = \delta_{ij}. \quad (2.496)$$

The diagonal currents ($j = 1, 2, \dots, N$)

$$: \lambda^{e_j} \lambda^{-e_j} : = : i \psi^{2j-1} \psi^{2j} : \equiv i J^{2j-1, 2j} \quad (2.497)$$

correspond to the generators of the Cartan subalgebra $\mathfrak{h} \subset \mathfrak{so}(2N)$. The remaining

$$\dim \mathfrak{so}(2N) - r_{\mathfrak{so}(2N)} \equiv 2N(2N-1)/2 - N \equiv 2N(N-1) \quad (2.498)$$

chiral $SO(2N)$ currents correspond to the *roots* of $\mathfrak{so}(2N)$ ⁹²

$$\{ \pm e_j \pm e_k, j \neq k \} \equiv \Delta(\mathfrak{so}(2N)) \subset \mathbb{Z}^N \quad (2.499)$$

and hence have the form

$$: \lambda^{\pm e_j} \lambda^{\pm e_k}(z) : \quad j \neq k. \quad (2.500)$$

Note that the $\mathfrak{so}(2N)$ roots are precisely the elements of squared-length 2 in the standard $\mathbb{Z}^N$ lattice (2.495), (2.496).

For a fixed j , the pair $\lambda^{e_j}(z), \lambda^{-e_j}(z)$ form a fermionic b, c system of weight $\lambda = 1 - \lambda = \frac{1}{2}$, and may be bosonized using the rules developed in Sect. 2.5. However,

⁹¹ For a background about lattices, see **BOX 2.10**.

⁹² See, for example, [56], **PLANCHE IV**.

BOX 2.10 - Lattices: basic definitions

We shall return to the general theory of *lattices* [72, 73] in Chap. 7. Here we limit to give the very basic definitions that are used in the main text.

Definition 2.1 A *lattice* is a pair $(\Lambda, \cdot)$ where Λ is a finitely generated, free, Abelian group and $\cdot : \Lambda \times \Lambda \rightarrow \mathbb{Z}$ is a non-degenerate, symmetric, integral, bilinear pairing.

A choice of *free generators* $\{e_1, \dots, e_r\} \subset \Lambda$ yields an isomorphism of Abelian groups

$$\mathbb{Z}^r \xrightarrow{\sim} \Lambda, \quad \mathbb{Z}^r \ni \mathbf{n} \equiv (n_1, n_2, \dots, n_r) \mapsto \sum_{i=1}^r n_i e_i \in \Lambda. \quad \clubsuit$$

The integer r is the *rank* of the lattice. Under the isomorphism ($\clubsuit$) the pairing $\cdot$ becomes

$$\mathbf{n}, \mathbf{n}' \mapsto \mathbf{n}' M \mathbf{n} \quad \text{where} \quad M_{ij} = e_i \cdot e_j \in \mathbb{Z}.$$

Two sets of generators $\{e_i\}$ and $\{e'_i\}$ differ by the action of the group $GL(r, \mathbb{Z})$. Then the lattice is specified (modulo isomorphism) by the rank r and the class $[M] \in S(r, \mathbb{Z})/GL(r, \mathbb{Z})$ of the Gram matrix M , where $S(r, \mathbb{Z})$ is the set of integral symmetric matrices M with $\det M \neq 0$ on which $GL(r, \mathbb{Z})$ acts as $M \rightarrow A' M A$ for $A \in GL(r, \mathbb{Z})$.

A lattice Λ is the *direct sum* of lattices, $\Lambda = \Lambda' \oplus \Lambda''$, iff M is in the $GL(r, \mathbb{Z})$ -class of the block diagonal matrix with blocks M' and M'' .

A lattice is called of *even* type if the diagonal elements $M_{ii} \in 2\mathbb{Z}$; in an even lattice $\lambda \cdot \lambda \in 2\mathbb{Z}$ for all $\lambda \in \Lambda$. A lattice which is not *even* is called of *odd* type.

The *dual lattice* $\Lambda^\vee$ of the lattice Λ is defined to be

$$\Lambda^\vee \stackrel{\text{def}}{=} \{\lambda \in \Lambda \otimes \mathbb{Q} : \lambda \cdot \kappa \text{ for all } \kappa \in \Lambda\}.$$

By definition $\Lambda \subset \Lambda^\vee$ is a sublattice of index

$$|\Lambda^\vee / \Lambda| = |\det M|.$$

A lattice is said to be *self-dual* iff $\Lambda^\vee \equiv \Lambda$, or, equivalently, if $\det M = \pm 1$.

A lattice is said to have *signature* (s, t) ($s + t = r$) if the corresponding bilinear product on the $\mathbb{R}$ -space $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ has signature (s, t) , i.e. if the symmetric matrix M has s positive and t negative eigenvalues. The lattice is called *definite* if $st = 0$ and *indefinite* otherwise.

The classification of lattice (up to isomorphism) will be given in Chap. 7.

Examples of positive-definite lattices are

- I_r , the lattice $\mathbb{Z}^r$ with Gram matrix $e_i \cdot e_j = \delta_{ij}$
- the root lattice of simply laced Lie algebras of type A_r , D_r and E_6 , E_7 , E_8 where the Gram matrix is the Cartan matrix C_{ij} of the Lie algebra.

The Witt theorem says that a positive-definite lattice generated by elements of length-square 1 or 2 is a direct sum of copies of I_r and simply laced root lattices.

the fact that the Hermitian conditions are *opposite*⁹³ to the ones used in Sect. 2.5 introduces some extra i 's in the formulae. We bosonize their Cartan currents as⁹⁴

$$:\lambda^{e_j}\lambda^{-e_j}(z): = i\partial\phi_j(z), \quad (2.501)$$

$$\phi_i(z)\phi_j(w) \sim -\delta_{ij}\log(z-w). \quad (2.502)$$

The usual bosonization rule would give

$$\lambda^{e_j} \stackrel{?}{=} e^{i\phi_j}, \quad \lambda^{-e_j} \stackrel{?}{=} e^{-i\phi_j}. \quad (2.503)$$

However, when $N > 1$ this is not totally correct since for $j \neq k$ the operators $e^{i\phi_j}$ and $e^{i\phi_k}$ commute rather than *anticommute* as they should in order to be identified with genuine Fermi fields. The Frenkel–Kač–Segal construction remedies this discrepancy by multiplying the exponential in the RHS of (2.503) by a *cocycle* which produces the missing signs.

In the $SO(2N)$ case the cocycles may be chosen⁹⁵ to coincide with the *Jordan–Wigner factors* [74] (i.e. with the recursive construction of fermionic operators or, equivalently, higher dimensional Dirac matrices [71]) that we are going to review.

Define the j th Fermi number

$$N_j \stackrel{\text{def}}{=} \frac{i}{2\pi} \oint dz \partial\phi_j \Rightarrow [N_j, e^{\pm i\phi_j}] = \pm e^{\pm i\phi_j}. \quad (2.504)$$

We refine Eq. (2.503) in the form

$$\lambda^{\pm e_i}(z) = c_{\pm e_j} e^{\pm i\phi_j(z)}, \quad (2.505)$$

$$\text{where } c_{\pm e_j} = \exp \left\{ \pm i\pi \left[\sum_{\ell < j} N_\ell \right] \right\}. \quad (2.506)$$

The symbol $[a]$ in (2.506) stands for the integer part of the real number a ; this peculiar prescription is required because of the phenomenon of Fermi number fractionalization:⁹⁶ the eigenvalues of the Fermi number N_ℓ in the Ramond sector are rational numbers not integers. With the prescription of subtracting the fractional part out of the sum of Fermi numbers in the exponential (2.506), we ensure that the *cocycle* operators $c_{\pm e_j}$ take the values ± 1 thus producing the extra signs needed for the operators in the RHS of (2.506) to anticommute in the appropriate way.

⁹³ In the usual quantization of b, c systems, say for the reparametrization ghosts, the fields b, c are taken to be Hermitian. Instead from the fact that ψ^μ is Majorana and Eq. (2.494) it follows that $(\lambda^{e_j})^\dagger = \lambda^{-e_j}$. Note that the Hermitian conditions on the ghosts b, c are required in order for the BRST charge Q_{BRST} to be Hermitian.

⁹⁴ Here ϕ_i is a chiral scalar, i.e. purely left-moving.

⁹⁵ The cocycle is not unique; see below.

⁹⁶ See Eq. (2.512) below for fractional Fermi number assignments of the R ground states.

More generally, let

$$\gamma = \sum_j n_j e_j \quad (2.507)$$

be an element of the Fermi charge lattice (2.495). To γ we associate the conformal primary operator

$$O_\gamma(z) \stackrel{\text{def}}{=} c_\gamma e^{i\gamma \cdot \phi(z)}, \quad \text{where } \gamma \cdot \phi(z) = \sum_i n_i \phi_i(z), \quad c_\gamma \equiv \prod_j (c_{e_j})^{n_j}. \quad (2.508)$$

The cocycles c_γ are *not* unique: different choices may be convenient for different applications. To extend the bosonization procedure to spin fields, it is convenient to choose cocycles of the form

$$c_\gamma = \exp(i\pi \gamma \cdot T \cdot N), \quad (2.509)$$

where T is a lower triangular $N \times N$ matrix with zeros on the diagonal and non-zero elements equal to ± 1 . Then we have the general operator identity

$$O_\gamma(z) O_{\gamma'}(w) = (z - w)^{\gamma \cdot \gamma'} e^{i\pi \gamma \cdot T \cdot \gamma'} e^{i\gamma \cdot \phi(z) + i\gamma' \cdot \phi(w)} c_{\gamma + \gamma'}. \quad (2.510)$$

Taking $\gamma' = -\gamma$ we conclude that $O_\gamma(z)$ is a Virasoro primary of weight

$$h = \frac{1}{2} \gamma \cdot \gamma. \quad (2.511)$$

Bosonized Spin Fields

The bosonized form of the spin fields $S_\alpha(z)$ are obtained by considering operators $O_\gamma(z)$ of the form (2.508) but now with “fractional” Fermi-charge vector γ

$$\gamma = (\pm \tfrac{1}{2}, \pm \tfrac{1}{2}, \pm \tfrac{1}{2}, \dots, \pm \tfrac{1}{2}) \in \tfrac{1}{2} \mathbb{Z}^N \quad (2.512)$$

given by a *weight* of the spinor representation under the maximal torus $U(1)^N \subset \text{Spin}(2N)$.⁹⁷ There are

$$2^N = 2^{[2N/2]} \quad (2.513)$$

weights of the form (2.512); this matches the number of components of a spinor in $d = 2N$. A $\text{Spin}(2N)$ -weight γ of the form (2.512) corresponds to a chirality $+1$ (resp. -1) spinor iff the number of $-\frac{1}{2}$ is *even* (resp. *odd*).

It is easy to check that the operators $O_\gamma(z)$ with $\text{Spin}(2N)$ -weight γ as in (2.512) obey the OPEs of the spin field $S_\alpha(z)$ in Eqs. (2.490)–(2.493): different choices of the cocycle c_γ lead to different representations of the Dirac γ -matrices. If $d = 2N$

⁹⁷ By $\text{Spin}(2N)$ we mean the simply connected Lie group with Lie algebra $\mathfrak{so}(2N)$ (D_N in Cartan’s notation). $\text{Spin}(2N)$ is a double cover of $SO(2N)$.

is a dimension where (Euclidean) Majorana spinor exist, and one wishes to get the current algebra written in a *Majorana representation* of the Dirac matrices, one must be a little careful with the choice of cocycles. See, for example, [75] for explicit cocycles producing convenient Majorana representations.

Lattices and Local Projections

There are various lattices in the game. First we have the lattice of $SO(2N)$ *weights* (as contrasted to the lattice of $\text{Spin}(2N)$ weights!) which is $\mathbb{Z}^N$ with the standard (positive-definite) inner product

$$\gamma \cdot \gamma' = \sum_{i=1}^N \gamma_i \gamma'_i; \quad (2.514)$$

see (2.495), (2.496). We have the *root sublattice* $\Gamma_{\text{root}} \subset \mathbb{Z}^N$, generated by the roots⁹⁸

$$\alpha \in \Delta(\mathfrak{so}(2N)) \quad (2.515)$$

of $\mathfrak{so}(2N)$, which are the elements of length-square 2 endowed with the induced inner product. In fact

$$\Gamma_{\text{root}} = \{\gamma \in \mathbb{Z}^N \mid \iota \cdot \gamma \in 2\mathbb{Z}\} \quad \text{where } \iota = (1, 1, \dots, 1). \quad (2.516)$$

Then we have the *dual* lattice $\Gamma_{\text{weight}} \equiv \Gamma_{\text{root}}^\vee$ defined by⁹⁹

$$\Gamma_{\text{weight}} \stackrel{\text{def}}{=} \{\gamma \in \mathbb{Q}^N \mid \gamma \cdot \alpha \in \mathbb{Z}, \forall \alpha \in \Gamma_{\text{root}}\}, \quad (2.517)$$

whose elements are the $\text{Spin}(2N)$ weights. Since $\alpha = \pm e_i \pm e_j$ ($i \neq j$) we have

$$\Gamma_{\text{weight}} = \left\{ (\gamma_1, \gamma_2, \dots, \gamma_N) \in \left(\frac{1}{2}\mathbb{Z}\right)^N \mid \gamma_i - \gamma_j = 0 \pmod{1} \right\}. \quad (2.518)$$

By general Lie theory [56]

$$\Gamma_{\text{weight}} / \Gamma_{\text{root}} \cong Z(\text{Spin}(2N)) = \begin{cases} \mathbb{Z}_2 \times \mathbb{Z}_2 & N \text{ even} \\ \mathbb{Z}_4 & N \text{ odd.} \end{cases} \quad (2.519)$$

$N = 4n$: Locality Structure For reasons to become clear in a moment, we focus on the case $N = 4n$, i.e. on the current algebra of the Lie group $\text{Spin}(8n)$, so

$$Z(\text{Spin}(8n)) = \mathbb{Z}_2 \times \mathbb{Z}_2. \quad (2.520)$$

⁹⁸ As already mentioned, the roots correspond to the $SO(2N)$ currents which are not in the Cartan subalgebra; cf. Eq. (2.499).

⁹⁹ For the relevant definitions of lattice theory, see **BOX 2.10**.

We note in passing that all the groups $\text{Spin}(8n)$ are related by *Bott periodicity*, that is, the corresponding even subalgebras of the universal *real* Clifford algebras are all Morita equivalent to the $\mathbb{R}$ -algebra $\mathbb{R} \oplus \mathbb{R}$ [71]. More generally [71] the Morita-equivalence with $\mathbb{R} \oplus \mathbb{R}$ extends to the Lorentz groups of signature (say) $(1, 1)$, $(25, 1)$, and $(9, 1)$, i.e. to the string world-sheet and the string/superstring target spaces in physical (Minkowskian) signature.

A nice property of the $\text{Spin}(8n)$ current algebras is that all local operators

$$O_\gamma(z) = c_\gamma e^{i\gamma \cdot \phi(z)}, \quad \gamma \in \Gamma_{\text{weight}} \quad (2.521)$$

are *local* with respect to themselves and their Hermitian conjugates $O_{-\gamma}(z)$, since

$$\gamma \cdot \gamma \in \mathbb{Z} \quad \text{for all } \gamma \in \Gamma_{\text{weight}}. \quad (2.522)$$

By Eq. (2.520) the quotient $\Gamma_{\text{weight}}/\Gamma_{\text{root}}$ consists of the four classes¹⁰⁰

$$\begin{aligned} (o) &= [(0, 0, \dots, 0, 0)], & (v) &= [(1, 0, \dots, 0, 0)], \\ (s) &= [(\tfrac{1}{2}, \tfrac{1}{2}, \dots, \tfrac{1}{2}, \tfrac{1}{2})], & (c) &= [(\tfrac{1}{2}, \tfrac{1}{2}, \dots, \tfrac{1}{2}, -\tfrac{1}{2})]. \end{aligned} \quad (2.523)$$

One has

$$(v) \cdot (s) \in \mathbb{Z} + \tfrac{1}{2}, \quad (v) \cdot (c) \in \mathbb{Z} + \tfrac{1}{2}, \quad (s) \cdot (c) \in \mathbb{Z} + \tfrac{1}{2}. \quad (2.524)$$

In order to get a physically sound $\text{Spin}(8n)$ -covariant, (chiral and unital) local operator algebra of the form

$$\mathcal{A} = \left\{ O_\gamma(z) \mid \gamma \in \Gamma \subset \Gamma_{\text{weight}} \right\} \quad (2.525)$$

we need to satisfy three conditions:¹⁰¹

- (a) Γ should be a lattice with $\Gamma \supset \Gamma_{\text{root}}$;
- (b) all operators $O_\gamma(z)$ with $\gamma \in \Gamma$ are *mutually local*, i.e. $\gamma \cdot \gamma' \in \mathbb{Z}$ for all $\gamma, \gamma' \in \Gamma$, that is, Γ must be an *integral* lattice;
- (c) Γ must be *maximal* with respect to properties a), b), i.e. Γ should be *self-dual*

$$\Gamma_{\text{root}} \subset \Gamma \equiv \Gamma^\vee \subset \Gamma_{\text{root}}^\vee \equiv \Gamma_{\text{weight}}. \quad (2.526)$$

The first condition says that $\mathcal{A}$ is closed under the OPE product and contains the $SO(8n)$ currents. The last condition says that we must be careful not to forget any additional local operator the theory may have: we will see in Chap. 5 that this *strong* version of the locality requirement is equivalent to modular invariance provided all

¹⁰⁰ As always, $[a]$ stands for the equivalence class of the element a .

¹⁰¹ In Sect. 5.1 we shall return to these conditions from a more general and rigorous perspective.

operators in $\mathcal{A}$ have integral conformal spin. By *a)*, if $\gamma \in \Gamma$ then all elements of its class should also be contained in Γ , i.e.

$$\gamma \in \Gamma \implies [\gamma] \subset \Gamma. \quad (2.527)$$

Then, in view of Eq. (2.524), Γ is integral iff it consists of the class (o) and at most *one* out of the three classes (v) , (s) , or (c) . If Γ consists of (0) and precisely one other class, condition *c)* is automatically satisfied.

Then we have *three* possible local 2d CFT:

- the one based on $(o) + (v)$ which is just the “obvious” theory of $2d$ fermions ψ^μ ;
- the *two* models with spin fields of definite chirality: $(o) + (s)$ and $(o) + (c)$.

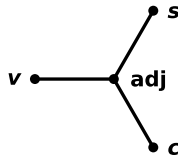
The last two local theories correspond to the two GSO projections which are essentially equivalent: indeed, flipping the sign of the scalar ϕ_1 , we flip $(s) \leftrightarrow (c)$.¹⁰²

We see that out of the set of all NS plus R operators—which is $\mathcal{O}_\gamma(z)$ with $\gamma \in \Gamma_{\text{weight}}$ —we projected out the NS operators with vector weights (v) and the R operators corresponding to one of the two chiralities. This corresponds to the projection $\Gamma_{d+1} \equiv (-1)^F = +1$ described around Eq. (2.487). Indeed NS states obtained by acting on the vacuum $|0\rangle$ with an odd number of fermions ψ^μ have weight in (v) and $(-1)^F = -1$, while states in (s) and (c) have opposite $(-1)^F$ eigenvalue: Γ_{d+1} is the $\text{Spin}(8n)$ chirality operator.

2.9.3 Spin(8) Triality and Refermionization

The $\text{Spin}(8)$ current algebra at level 1 has special properties which are crucial in superstring theory both conceptually and as computational devices.¹⁰³

The Dynkin graph of the Lie algebra $\mathfrak{spin}(8) \equiv D_4$



(2.528)

has a $\mathfrak{S}_3$ group of automorphisms which permutes the three peripheral nodes. This symmetry induces a group of outer automorphism of the Lie algebra called *triality* [77]. The fundamental representations associated with the three peripheral nodes

¹⁰² In target-space language: spacetime parity flips the chirality of fermions.

¹⁰³ $\text{Spin}(8)$ is the transverse subgroup of the 10d Lorentz group $SO(9, 1)$ which is manifest in the unitary light-cone quantization of the superstring. Refermionization then relates the NS-R formulation of the superstring to the GS one in the light-cone gauge [76].

[54, 55] are the vector v , and the two spinors, s and c , of opposite chirality: all three representations are real of dimension 8. Triality permutes the representations v, s, c and fixes the adjoint representation associated with the central node in (2.528). Then

$$\mathfrak{spin}(8) \simeq v \wedge v \simeq s \wedge s \simeq c \wedge c. \quad (2.529)$$

By **Rule 2.1** at $k = 1$ the non-trivial primary operators $\lambda^i(z)$, $S_\alpha(z)$, and $S_{\tilde{\alpha}}(z)$ transform in the representations v, s , and c , respectively. We bosonize the level 1 $\mathbf{Spin}(8)$ current algebra in terms of four free chiral scalars ϕ_a as in Sect. 2.9.2. The primaries $\psi^i(z)$ in representation v have the form ($a, b = 1, \dots, 4$)

$$\begin{aligned} \psi^{2a-1}(z) \pm i\psi^{2a}(z) &= c_{\pm e_a} \exp(\pm i\phi_a(z)), \\ \text{with } \phi_a(z)\phi_b(w) &\sim -\delta_{ab} \log(z-w), \end{aligned} \quad (2.530)$$

and have weights $(\frac{1}{2}, 0)$ and so are free fermions. The spin fields $S^\alpha(z)$ in representation s can be written in the form¹⁰⁴

$$\begin{aligned} S^{2a-1}(z) \pm iS^{2a}(z) &= c_{\pm e'_a} \exp(\pm i\phi'_a(z)), \\ \text{with } \phi'_a(z)\phi'_b(w) &\sim -\delta_{ab} \log(z-w), \end{aligned} \quad (2.531)$$

where the canonical free fields $\phi'_a \equiv S_{ab}\phi_b$ are linear combinations of the ϕ_a 's:

$$\begin{bmatrix} \phi'_1 \\ \phi'_2 \\ \phi'_3 \\ \phi'_4 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix}, \quad S \equiv \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}. \quad (2.532)$$

We see from these formulae that the $S_{\pm a}(z)$ are also primaries of weights $(\frac{1}{2}, 0)$ hence free fermions. From (2.529) one gets

$$\psi^i \psi^j(z) = \frac{1}{2} S^\alpha (\Gamma^{ij})_{\alpha\beta} S^\beta(z) = \frac{1}{2} S^{\dot{\alpha}} (\Gamma^{ij})_{\dot{\alpha}\dot{\beta}} S^{\dot{\beta}}(z) \quad (2.533)$$

where Γ^{ij} are $\mathbf{Spin}(8)$ Dirac matrices and α, β (resp. $\dot{\alpha}, \dot{\beta}$) are chirality + (resp. -) Weyl spinor indices. We conclude that the current algebra $\mathbf{Spin}(8)_1$ can be written in terms of *three different* (but isomorphic) sets of free fermions. The three possible “GSO projections”, i.e. maximal local subalgebras, are permuted by triality. The triality automorphism interchanges the vector representation and the two spinor representations. Triality is equivalent to the following isomorphism of the three maximally local “GSO” lattices (see Sect. 2.9.2)

$$(o) + (v) \simeq (o) + (s) \simeq (0) + (c) \quad (2.534)$$

¹⁰⁴ The primaries in c are obtained from the ones in s by making $\phi_1 \leftrightarrow -\phi_1$.

where the first isomorphism is given by the matrix S_{ab} in Eq. (2.533). In the bosonization of the level-1 $\mathfrak{spin}(8)$ current algebra, we associated with each maximally local ($\equiv$ self-dual) lattice $\Gamma \simeq \Gamma^\vee \subset \Gamma_{\text{weight}}$ a complete local operator algebra $\mathfrak{A}(\Gamma)$

$$\mathfrak{A}(\Gamma) \stackrel{\text{def}}{=} \{c_\gamma e^{i\gamma \cdot \phi(z)} : \gamma \in \Gamma\} \quad (2.535)$$

where the one associated with $\Gamma = (0) + (v)$ contains the original fermions ψ^i and no spin fields, while $\Gamma = (0) + (s)$ and $\Gamma = (0) + (c)$ contain only NS operators with an even number of Fermi fields and spin fields of positive, resp. negative, chirality. Triality says that these three operator algebras—in the special case of $\mathfrak{spin}(8)$ —are actually isomorphic under the map

$$c_\gamma e^{i\gamma \cdot \phi(z)} \rightarrow c_{(S\gamma)} e^{i(S\gamma) \cdot \phi(z)}. \quad (2.536)$$

The chain of changes of d.o.f.

$$\psi^i \rightsquigarrow \phi_a \rightsquigarrow \phi'_a \rightsquigarrow S^\alpha, \quad (2.537)$$

which starts from fermions ψ^i , goes through bosons ϕ^a , and ends up again with fermions S^α , is called *refermionization*. By the very notion of spin fields, the final fermions S^α are non-local with respect to the original ones ψ^i .

To define the fermion field ψ^i (resp. S^α) on a world-sheet Σ , we need to fix a spin-structure $\mathcal{L}$ (resp. $\mathcal{L}'$). An amplitude written in terms of the free fermion S^α in a spin-structure $\mathcal{L}'$ is a sum over *all* possible spin-structures of amplitudes written in terms the fermion ψ^i of the form

$$\left\langle \cdots \cdots \right\rangle_{S^\alpha, \mathcal{L}'} = \sum_{\mathcal{L}} c(\mathcal{L}', \mathcal{L}) \left\langle \cdots \cdots \right\rangle_{\psi^i, \mathcal{L}} \quad (2.538)$$

for some constants $c(\mathcal{L}', \mathcal{L})$. We shall use the refermionization identities (2.538) to dramatically simplify computations. These identities lead to relations between theta-functions such as the Jacobi and Riemann ones [78]. We shall return to refermionization in Sects. 6.3 and 10.3 where we determine the constants $c(\mathcal{L}', \mathcal{L})$.

2.10 On Classification of 2d Superconformal Algebras

For the sake of completeness we make a rapid survey of the other superconformal algebras in 2d and explain why they are less relevant for string theory.

Chiral and Supersymmetry Algebras

From the point of view of the world-sheet theory, both the bosonic string and the superstring are defined in terms of an algebra of constraints which in modern times we quantize as *à la* BRST. This algebra is generated by the modes of a set of holomorphic

and anti-holomorphic currents of various spins: the energy–momentum tensor $T_B(z)$ in the bosonic case, and the energy–momentum tensor together with its fermionic superpartner $T_F(z)$ (the $\mathcal{N} = 1$ supercurrent) in the superstring case.

Looking for a more systematic treatment, we ask for the classification of all the possible symmetry algebras which may be used as an algebra of constraints to define the world-sheet theory of a meaningful string theory. Of course, the very same symmetry algebras may be also realized in the 2d theory as *global* symmetries rather than *gauge* symmetries. In this case the physical Hilbert space carries a non-trivial representation of the symmetry algebra, which relates amplitudes and observables for physically distinct states. CFTs with such global extended symmetry algebras (and the appropriate central charge) may be seen as non-trivial string backgrounds having special properties: see Chap. 11 for this important application of the extended chiral superalgebras. For the moment we are concerned with the allowed algebras of gauge constraints. We focus on the *chiral algebras*, i.e. the algebras generated by holomorphic currents.

Our theory should be invariant under 2d reparametrizations, so we require the energy–momentum $T(z)$ to be part of $\mathcal{A}$. Then $\mathcal{A}$ contains Virasoro as a subalgebra. We assume that the induced Virasoro representation on $\mathcal{A}$ is unitary¹⁰⁵; then the several holomorphic currents have CFT weights of the form $(h, 0)$ with $h \geq 0$ while $2h \in \mathbb{Z}$ by topological considerations. Since $\tilde{h} = 0$, h is also the spin of the chiral current. Without loss of generality, we may assume the currents to be Hermitian by taking their real and imaginary parts. We consider the various possibilities in turn.

The algebra $\mathcal{A}$ contains currents with $h > 2$ Chiral algebras with $h > 2$ currents are called *W-algebras* [79]. Many of them are known, but there is no complete classification. Sometimes they appear in string models as global symmetry algebras.

There have been attempts to use them as algebras of constraints [80, 81], but the fact that the commutator of two generators is non-linear in the generators makes the BRST quantization quite subtle. The few examples of *W*-strings so constructed appear to be equivalent to theories which can be constructed using more standard chiral algebras of constraints. In addition, we do not want to lose the geometric picture of the world-sheet as the world-story of the string. For the *W*-strings the geometrical interpretation is not very clear (and would certainly imply some extra geometric structure on the world-sheet, which will make its geometry “less intrinsic”). For these reasons we restrict to chiral algebras with $h \leq 2$ currents.

The algebra $\mathcal{A}$ contains several $h = 2$ currents The world-sheet theory may indeed contain several $h = 2$ chiral currents. For instance, in the closed superstring with the matter action in Eq. (2.397), each free field X^μ , ψ^ν has its own conserved energy–momentum tensor. However, only the *total* energy–momentum tensor enters in the constraint algebra; this follows from the geometric interpretation of the constraint algebra as implementing the world-sheet $\mathbf{Diff}^+$ invariance. The spin-2 constraint

¹⁰⁵ This condition does not hold in string theory because of the ghosts and the time-like fields. However, the matter sector when Wick rotated to spacetime Euclidean signature is expected to be unitary as a 2d theory, and we restrict ourselves to this situation. Yet there are interesting string theories whose 2d models are non-unitary; see the **Remark** at the end of this section.

current is the one dual (i.e. coupled) to the (unique) world-sheet metric g_{ab} . Then we assume that $\mathcal{A}$ contains a unique $h = 2$ holomorphic current, which is the total energy–momentum tensor, generating the Virasoro subalgebra contained in $\mathcal{A}$.

Chiral algebras containing $h \notin \frac{1}{2}\mathbb{Z}$ currents If the spin h of a holomorphic current $j(z)$ is not integral or half-integral, the current is not local with respect to itself since the OPE

$$j(z) j(w) \sim \frac{C}{(z-w)^{2h}} \quad (2.539)$$

is multivalued. There are many CFTs with such *fractional* spin currents as global symmetries. However, their non-locality introduces severe complications if one tries to use them as gauge constraints. It is not clear if such *fractional strings* may be consistently defined (see [82] for a survey of attempts).

In view of these considerations, we restrict to algebras satisfying the following:

Assumption 2.1 We assume that the constraint algebra $\mathcal{A}$ is a *local* chiral algebra, i.e. an algebra generated by a finite set of mutually local currents, consisting of a unique spin 2 (Virasoro) current $T(z)$ together with currents of spin $\frac{3}{2}$, 1, $\frac{1}{2}$, and 0. $\mathcal{A}$ is required to carry an unitary representation of its Virasoro subalgebra. In particular, we assume the 2d Spin and Statistics theorem to hold, so that currents of integer (half-integer) h commute (resp. anticommute). Finally, to avoid the non-linearities mentioned above, we assume the modes of the currents of $\mathcal{A}$ to generate a (centrally extended) Lie superalgebra, i.e. the singular part of the OPE of two currents of $\mathcal{A}$ is at most linear in the currents of $\mathcal{A}$.

Faddeev–Popov Ghosts for $\mathcal{A}$

Under **Assumption 2.1**, the BRST quantization of the theory with constraint algebra $\mathcal{A}$ associates with each current $J^h(z) \in \mathcal{A}$ of Virasoro weight h integral (resp. half-integral) an anticommuting b, c ghost system (resp. a commuting β, γ ghost system) with $\lambda = h$ as discussed in Sect. 2.5. From that section we know that the central charge c_h of each ghost system is (cf. Eq. (2.236))

$$c_h = \epsilon(1 - 3Q^2), \quad \text{where} \quad \begin{cases} \epsilon = (-1)^{2h+1}, \\ Q = \epsilon(1 - 2h), \end{cases} \quad (2.540)$$

that is,

	$h = 2$	$h = \frac{3}{2}$	$h = 1$	$h = \frac{1}{2}$	$h = 0$	
ϵ	+1	−1	+1	−1	+1	
Q	−3	+2	−1	0	+1	
c_h	−26	+11	−2	−1	−2	(2.541)

As we shall see in Chap. 3, the consistency of BRST quantization requires all *total* central extensions of $\mathcal{A}$, and in particular the total Virasoro central charge c_{tot} , to vanish

$$c_{\text{tot}} \equiv c_{\text{matter}} + c_{\text{ghosts}} = 0. \quad (2.542)$$

From the table (2.541) we see that only the ghosts of the $\text{spin}-\frac{3}{2}$ currents give a *positive* contribution to the total (“matter” + ghosts) Virasoro central charge c_{tot} . This simple observation has interesting implications for our discussion.

Algebras $\mathcal{A}$ with No $\text{Spin}-\frac{3}{2}$ Currents

Let $\mathcal{A}$ satisfy **Assumption 2.1** and have no $\text{spin}-\frac{3}{2}$ current. From Eq. (2.541) we see that $c_{\text{ghost}} \leq -26$ with equality iff $\mathcal{A}$ is generated by $T(z)$, i.e. iff $\mathcal{A}$ is the Virasoro algebra, in which case the constrained system reduces to the bosonic string.

The general case consists of a matter sector which contains additional degrees of freedom with respect to the bosonic string (since $c_{\text{matter}} > 26$), but the additional constraints precisely remove the added matter states since the CFT of “matter plus the $h \leq 1$ ghost systems” is equal to $+26$. Thus the theory is just the bosonic string in which the matter is a certain CFT with $c = 26$ which should be unitary for the theory to be ghost-free. Thus we do not get an essentially new theory, but only a variant of the bosonic string in which the usual free $c = 26$ “matter” CFT is replaced by a more general one. This “matter” CFT has a deflist as a “gauge” theory, corresponding to its $h \leq 1$ constraint algebra. It may be that this gauge approach is a convenient way to construct/describe the given $c = 26$ CFT, so the formulation with extra $h \leq 1$ ghost may be useful in many contexts. Nevertheless an algebra of chiral constraints without $\text{spin}-\frac{3}{2}$ currents does *not* lead to an essentially new class of string theories, and hence is not relevant for our present classification purposes.

Thus, for our present goals, we may limit ourselves, in addition to the Virasoro algebra (bosonic string), to chiral algebras $\mathcal{A}$ satisfying **Assumption 2.1** and having $N \geq 1$ $\text{spin}-\frac{3}{2}$ holomorphic currents $T_F^a(z)$.

Since the charges associated with $\text{spin}-\frac{3}{2}$ currents are SUSY supercharges, the interesting chiral algebras $\mathcal{A}$ are just the $2d$ (*chiral*) *supersymmetry algebras*, possibly extended by additional holomorphic currents. The chiral currents should form complete SUSY supermultiplets. We conclude this general discussion with a couple of general remarks.

Note 2.10 A non-degenerate unitary CFT has only one operator with $h = \tilde{h} = 0$, namely the identity. Thus the possibility of $h = 0$ chiral currents in $\mathcal{A}$ should be taken with a *pinch of salt*. The standard interpretation is to trade the scalar $h = 0$ current $\chi(z)$ for the $h = 1$ current $\partial\chi(z)$. As in the case of the scalar ghost field $\xi(z)$ (cf. Sect. 2.5.6) this replacement kills the constant zero-mode of $\chi(z)$.

Remark 2.1 In the **Assumptions 2.1**, we may give up the unitarity requirement and, in particular, the *Spin and Statistics* connection. If we relax the **Assumptions** new possibilities open up. The most interesting one is the *topological string theory* which is essentially solvable and has deep mathematical applications. For some background see, for example, [13].

2d Superconformal Algebras

The chiral supersymmetry algebras, containing the Virasoro algebra, are called *superconformal* algebras. Such an algebra $\mathcal{A}$ contains

Table 2.1 Superconformal algebras (including the bosonic one)

$\mathcal{N}$	n_1	$n_{1/2}$	n_0	c_{ghost}	H	R	$\mathfrak{f}$
0	0	0	0	-26			$\mathfrak{sl}(2)$
1	0	0	0	-15			$\mathfrak{osp}(1 2)$
2	1	0	0	-6	$U(1)$	± 1	$\mathfrak{osp}(2 2)$
3	3	1	0	0	$SU(2)$	3	$\mathfrak{osp}(3 2)$
4	3	0	0	12	$SU(2)$	$2 \oplus 2$	$\mathfrak{su}(2 1, 1)$
4	6	4	1	0	$SU(2) \times SU(2)$	(2,2)	$\mathfrak{osp}(4 2)$
4	7	4	0	0	$SU(2) \times SU(2) \times U(1)$	(2,2,0)	$D(2 1; \alpha) \oplus \mathfrak{u}(1)$

- a unique energy–momentum tensor $T(z)$ with $h = 2$;
- a number of Hermitian anticommuting spin $\frac{3}{2}$ chiral supercurrents $G^a(z)$, $a = 1, 2, \dots, \mathcal{N} \geq 1$. These currents are Virasoro primaries with $h = 3/2$

$$T(z) G^a(w) \sim \frac{3}{2} \frac{G^a(w)}{(z-w)^2} + \frac{\partial G^a(w)}{z-w}, \quad (2.543)$$

and their NS modes $G_{1/2}^a$ decrease the weight of a state by $1/2$;

- (possibly) conserved $h = 1$ currents $J^k(z)$ ($k = 1, \dots, d \geq 0$) whose associated charges generate a Lie group H . The $J^k(z)$ generate the affine Lie algebra $H^{(1)}$ and have OPE

$$J^k(z) J^h(w) \sim \frac{K^{kh}}{(z-w)^2} + \frac{f^{kh}_j J^j(w)}{z-w}, \quad (2.544)$$

with K^{kh} a positive-definite H -invariant symmetric tensor. The supercurrents $G^a(z)$ transform in some unitary representation R of H and hence

$$J^k(z) G^a(w) \sim \frac{R^{ka}_b G^b(w)}{z-w} \quad (2.545)$$

- (possibly) spin 0 and spin $\frac{1}{2}$ currents.

We summarize the classification of 2d superconformal algebras in Table 2.1. The interested reader may find detailed proofs of the classification in Sect. 2.10.1. We conclude this section by outlining the implications of Table 2.1 for string theory.

Implications for Stringy Constructions

We were led to the classification in the Table 2.1 by the search for suitable world-sheet constraint algebras $\mathcal{A}$. As in the bosonic and fermionic string, one starts with a “matter” 2d QFT in a unitary representation of the symmetry $\mathcal{A}$ and gauges this symmetry, that is, adds the Faddeev–Popov ghosts for all currents of $\mathcal{A}$, constructs

the associated BRST charge Q , and defines the physical states as the appropriate Q -cohomology. However, as we already know, a necessary condition for the nilpotency condition $Q^2 = 0$ is that

$$c_{\text{matter}} + c_{\text{ghosts}} = 0, \quad (2.546)$$

while $c_{\text{matter}} > 0$ for unitary “matter”. Therefore for all algebras $\mathcal{A}$ such that $c_{\text{ghost}} \geq 0$ there is no possible “matter” CFTs which satisfies Eq. (2.546), hence no meaningful stringy theory. Then, by a part for the bosonic string algebra $\mathcal{N} = 0$ and the $\mathcal{N} = 1$ superstring already introduced in Chap. 1, we have only a new candidate, i.e. $\mathcal{N} = 2$ which we shall discuss in the **Appendix** to Chap. 7. The $\mathcal{N} = 2$ string theory—while extremely interesting mathematically—is not relevant for the “real” physics since it predicts a signature of spacetime $(2, 2)$, that is, with *two* times; see Chap. 7.

Heterotic Constructions

There is, however, yet another possibility. $\mathcal{A}$ was a chiral algebra of left-moving (holomorphic) constraints; of course, there is also a right-moving such algebra $\tilde{\mathcal{A}}$. There is no reason for the two chiral algebras to be isomorphic, as they were in the bosonic and superstring theories introduced in Chap. 1. We may consider a pair of superconformal algebras $(\mathcal{A}, \tilde{\mathcal{A}})$ with $\mathcal{N} \neq \tilde{\mathcal{N}}$, of the kind $(0, 1)$ or $(1, 0)$. This asymmetric possibility leads to the *heterotic string* which is a major topic in this textbook (see Chap. 7). One may consider the cases $(\mathcal{N}, \tilde{\mathcal{N}}) = (2, 1)$ or $(2, 0)$ but they are less physically interesting for reasons explained in the **Appendix** to Chap. 7. One basic aspect should be kept in mind: heterotic strings, having $\mathcal{N} \neq \tilde{\mathcal{N}}$ local SUSY on the world-sheet, are necessarily *closed and oriented*, since boundary conditions and Ω projections identify left- and right-moving SCFTs which is possible only if the two theories are isomorphic.

2.10.1 ★ Classification of 2d Superconformal Algebras

In view of **Assumption 2.1**, the $G^a(z)$, being Grassmann odd, satisfy an OPE of the general form

$$G^a(z) G^b(w) \sim \frac{C^{ab}}{(z-w)^3} + \frac{2M^{ab}_k J^k(w)}{(z-w)^2} + \frac{2S^{ab} T(w) + M^{ab}_k \partial J^k(w)}{z-w} \quad (2.547)$$

for some constant matrices

$$C^{ab} = C^{ba}, \quad M^{ab}_k = -M^{ba}_k, \quad \text{and} \quad S^{ab} = S^{ba}. \quad (2.548)$$

Both C^{ab} and S^{ab} are positive-definite¹⁰⁶ H -invariant tensors in $\odot^2 R$.

Computing the three-point function

$$\langle G^a(z) G^b(w) T(y) \rangle_{S^2} \quad (2.549)$$

¹⁰⁶ From Eq. (2.547) it follows $\{G^a_{1/2}, G^b_{-1/2}\} = 2S^{ab}L_0 + M^{ab,k}J^k_0$ while $G^a_{1/2} = (G^a_{-1/2})^\dagger$. Then (in a unitary theory) S^{ab} is the matrix of a positive-definite inner product.

in two different ways (i.e. using first the $G^a G^b$ OPE or first the $G^a T$ OPE), we see that

$$C^{ab} = \frac{2c}{3} S^{ab}, \quad (2.550)$$

where c is the Virasoro central charge. We may set $S^{ab} = \delta^{ab}$ by going to an orthonormal basis for the supercurrents $G^a(z)$. In the same fashion from the three-point function

$$\langle G^a(z) G^b(w) J^k(y) \rangle_{S^2} \quad (2.551)$$

we get the equality

$$2M^{ab}{}_k K^{kh} = R^h{}_a C^{cb} \equiv \frac{2c}{3} R^h{}_a b. \quad (2.552)$$

We may set $K^{kh} = c \delta^{kh}/3$ as a normalization of the $h = 1$ currents. Then, with this normalization,

$$M^{ab}{}_k \equiv R^{ka}{}_b \quad (2.553)$$

(in an orthonormal basis we do not need to distinguish upper and lower indices). The result (2.553) says that the supercurrents $G^a(z)$ are charged only with respect to $h = 1$ currents appearing in the $G^a(z) G^b(w)$ OPE, and, vice versa, if a current $J^k(z)$ appears in this OPE, the supercurrents $G^a(z)$ should be charged with respect to the associated charge J^k_0 . We split the $h = 1$ currents in $\mathcal{A}$ in two sets: the set of currents entering in the $G^a(z) G^b(w)$ OPE, which we write in the form

$$J^{ab}(z) = -J^{ba}(z) \stackrel{\text{def}}{=} M^{ab}{}_k J^k(z), \quad (2.554)$$

and the complementary set of currents under which the $G^a(z)$'s are *inert*. Correspondingly,

$$H = H_R \times F, \quad (2.555)$$

where H_R is the R -symmetry group generated by the $J^{ab}(z)$ currents and F is the group which leaves the supercurrents invariant (the “flavor group” in the usual jargon). We stress that the currents $J^{ab}(z)$ are **not** necessarily non-zero.

We rewrite the OPEs in terms of the Lie superalgebra of the (super)current modes

$$T(z) = \sum_{n \in \mathbb{Z}} \frac{L_n}{z^{n+2}}, \quad G^a(z) = \sum_{r \in \mathbb{Z} + \nu} \frac{G^a_r}{z^{r+3/2}}, \quad J^{ab}(z) = \sum_n \frac{J^{ab}_n}{z^{n+1}}. \quad (2.556)$$

From Eq. (2.547) and the above considerations we get

$$\{G^a_r, G^b_s\} = 2\delta^{ab} L_{r+s} + (r-s) J^{ab}_{r+s} + \frac{c}{3} \left(\frac{r^2}{4} - 1 \right) \delta^{ab} \delta_{r+s,0}, \quad (2.557)$$

while from (2.553)

$$[G^a_m, G^c_r] = R^{ab}{}^{cd} G^d_{m+r} \equiv (\delta^{ac} \delta^{bd} - \delta^{ad} \delta^{bc}) G^d_{m+r} = \delta^{ac} G^b_{m+r} - \delta^{bc} G^a_{m+r}. \quad (2.558)$$

We conclude that $H_R \subset SO(N)$, while the representation R is the restriction to the subgroup H_R of the vector one. The full group $SO(N)$ acts on the Lie superalgebra (2.557), (2.558) by automorphisms: therefore the Lie algebra $\mathfrak{h}_R \equiv \mathfrak{Lie}(H_R)$ is an *invariant ideal* in $\mathfrak{so}(N)$. For $N \neq 4$, $\mathfrak{so}(N)$ is *simple* and H_R is either¹⁰⁷ trivial or $SO(N)$. In particular, we have

$$\{G^a_{1/2}, G^b_{1/2}\} = 2\delta^{ab} L_1, \quad [L_1, G^a_{1/2}] = 0. \quad (2.559)$$

¹⁰⁷ We identify Lie groups modulo isogeny.

Instead of studying the SUSY representation on the generators of $\mathcal{A}$, we may focus on the representation on the states which correspond to these generators under the CFT state-operator map. We start with the state $|T\rangle \equiv L_{-2}|0\rangle$ corresponding to the energy–momentum tensor $T(z)$. It has $h = 2$ and is **quasi**-primary

$$L_1|T\rangle = 0. \quad (2.560)$$

Equations (2.559), (2.560) imply that the state

$$|a_1, a_2, \dots, a_s\rangle \equiv G_{1/2}^{a_1} G_{1/2}^{a_2} \cdots G_{1/2}^{a_s} |T\rangle \quad \left[\begin{array}{l} \text{has } h = 2 - s/2 \\ \text{in a representation of} \\ H \text{ contained in } \wedge^s R. \end{array} \right. \quad (2.561)$$

This collection of states for $s = 0, 1, \dots, N$ forms a supermultiplet. This implies that the number of bosonic and fermionic states are equal. In particular, one has

$$|a\rangle \equiv G_{1/2}^a |T\rangle = G_{1/2}^a L_{-2}|0\rangle = [G_{1/2}^a, L_{-2}]|0\rangle = \frac{5}{2} G_{-3/2}^a |0\rangle \equiv \frac{5}{2} |G^a\rangle \neq 0, \quad (2.562)$$

$$|a, b\rangle = \frac{5}{2} G_{1/2}^a G_{-3/2}^b |0\rangle = \frac{5}{2} \{G_{1/2}^a, G_{-3/2}^b\} |0\rangle = 5 J_{-1}^{ab} |0\rangle \equiv 5 |J^{ab}\rangle. \quad (2.563)$$

Consider the case $N = 5$. We have either $H_R = 1$ or $SO(5)$. In the first case all states $|a, b\rangle = 0$ vanish, and hence all states (2.561) with $s > 1$ are zero. We are left with one bosonic state $|T\rangle$ and the 5 fermionic states $|G^a\rangle$, so we get a contradiction with supersymmetry. We conclude that $H_R = SO(5)$ and $R \equiv V$ is the vector representation. But $\wedge^s V$ are irreducible representations of $SO(5)$ [83], so for a given s either all states (2.561) vanish or all are non-zero. Moreover, if they vanish for a given s_0 they vanish for all $s \geq s_0$. Let $\Delta(s_0)$ be the difference in the number of non-zero Bose and Fermi states of the form (2.561) under the assumption that the smaller s for which the states vanish is s_0 . One has

$$\Delta(2) = -4, \quad \Delta(3) = +6, \quad \Delta(4) = -4, \quad \Delta(5) = 1. \quad (2.564)$$

Thus, to get equal number of states we need $|1, \dots, 5\rangle \neq 0$, but this is in contrast with unitarity since the last states have $h = -\frac{1}{2}$. We conclude that $N = 5$ is not allowed. Since all algebras with $N \geq 5$ have a subalgebra with $N = 5$, we learn that a unitary 2d SCFT has $N \leq 4$. Let us consider the remaining cases one by one. For $N = 1$ we get back the superconformal algebra considered in Sect. 2.8 (plus possibly F currents commuting with the supercharges and their SUSY partners). For $N = 2$ the counting argument gives $\Delta(2) = -1$ so that the $h = 1$ state

$$|J\rangle \equiv |J^{12}\rangle \quad (2.565)$$

must be *non-zero*. Thus for $N = 2$ we have $H_R = SO(2) \simeq U(1)$ and the two supercharges $G^\pm(z)$ carry charges ± 1 .

For $N = 3$, $\Delta(2) = -2$, then $H_R = SO(3) \simeq SU(2)$. All representations $\wedge^s V$ are irreducible, and the counting argument shows that both $|a, b\rangle$ and $|1, 2, 3\rangle$ are non-zero. We have three $h = 1$ and one $h = \frac{1}{2}$ current (i.e. a free Majorana–Weyl fermion).

$N = 4$ is the tricky case. We have two possible non-trivial normal subgroups H_R :

$$SO(4) \simeq SU(2) \times SU(2) \quad \text{or} \quad SU(2). \quad (2.566)$$

In the second case the R-currents should satisfy the self-dual constraint

$$J^{ab}(z) = \pm \frac{1}{2} \epsilon^{abcd} J^{cd}(z) \quad (2.567)$$

for one of the two signs $\pm$. This self-duality condition implies $|a, b, c\rangle = 0$, and we are left with the following currents: one $T(z)$, four $G^a(z)$, and three J^{ab} , giving $1 - 4 + 3 = 0$, as it should.

On the other hand when $H_R = SU(2) \times SU(2)$, all states (2.561) of given s should be present or absent; the only possibility of getting equal number of bosonic/fermionic states is that all states are non-zero. $|1, 2, 3, 4\rangle$ is a singlet current $\chi(z)$ which has weight 0 and which is naturally replaced by a singlet $h = 1$ current $J(z) = \partial\chi(z)$ whose charge leaves invariant the supercurrent, i.e. which belongs to F . The same argument shows that for $N \geq 2$ we cannot have further $h \leq 1$ currents (“flavor” currents) because the corresponding extended supermultiplets do not exist (at least under the assumption that the identity is the only operator with $h = \tilde{h} = 0$).¹⁰⁸

For $N = 1$, we may neglect the possibility of F currents by the same argument we neglected them (for the purposes of the present section!) in the bosonic string case.

In conclusion we are left with the superconformal algebras in Table 2.1, where n_h is the number of currents of weight h with $n_2 = 1$ and $n_{3/2} = N$; c_{ghost} is the total Virasoro central charge of the corresponding ghost system computed from equation (2.541). The last column gives the corresponding finite-dimensional Lie superalgebra¹⁰⁹ $\mathfrak{f}$ in the Kač notation [84]. Note that the superalgebra in the last row depends on a free real parameter $\alpha \neq 0, -1$; in fact, in the last case we have a *one parameter family* of non-isomorphic superconformal algebras $\mathcal{A}$. For details on this very special superalgebra, see BOX 2.11.

Appendix 1: Witten’s Non-abelian 2d Bosonization

The Abelian bosonization is not very convenient when we have a system with many fermions. For example, N free massless Majorana 2d fermions have $O(N) \times O(N)$ chiral symmetry; only the Cartan torus is a manifest symmetry in its Abelian bosonized form, and there is no nice expression for the currents associated with the roots of the $\mathfrak{o}(N)$ algebra. We wish to generalize bosonization to a fully non-Abelian scheme, where all the symmetries are manifest and all conserved currents have a simple (local) operator counterpart in the bosonic theory. This was achieved by Witten [85] and Polyakov–Wiegmann [86].

Computing the Path Integral We start by computing

$$\text{Det}[i\mathcal{D}] = \int [d\bar{\psi} d\psi] e^{-\int \bar{\psi} \gamma^\mu (i\partial_\mu + A_\mu) \psi d^2z} \quad (2.568)$$

or, rather, its Majorana form

$$\text{Pf}[i\mathcal{D}] = \int [d\psi] e^{-\int \bar{\psi} \gamma^\mu (i\partial_\mu + A_\mu) \psi d^2z} \quad (2.569)$$

¹⁰⁸ Let $|J\rangle \equiv J_{-1}|0\rangle$ be the state associated with an additional current $J(z)$ which commutes with H_R . The states $G_{1/2}^{a_1} \cdots G_{1/2}^{a_s}|J\rangle$, if non-zero, have $h = 1 - s/2$, and belong to the same H_R representations as the $|a_1, \dots, a_s\rangle$. The same counting argument, as before, implies that if $|J\rangle \neq 0$, all these states are non-zero for $N \neq 4$ while for $N = 4$ at least three states $G_{1/2}^{a_1} \cdots G_{1/2}^{a_s}|J\rangle$ are non-zero. So for $N \neq 2, 4$ we have states with $h < 0$ contrary to unitarity. For $N = 2$ (resp. 4) we have one (resp. three) non-trivial operator with $h = 0$.

¹⁰⁹ This superalgebra, $\mathfrak{f}$, is the maximal superconformal symmetry which may be linearly realized quantizing the theory on the sphere, that is, the largest Lie* sub(super)algebra of $\mathcal{A}$ which leaves the vacuum invariant $|0\rangle$. A current of $\mathcal{A}$ of weight h contributes $2h - 1$ generators to $\mathfrak{f}$ which then has $3 + n_1$ bosonic generators and $2N$ fermionic ones.

BOX 2.11 - The Lie superalgebra $D(2|1; \alpha)$

The Lie superalgebras $OSp(m|n)$ (n even) have bosonic subalgebra $\mathfrak{so}(m) \oplus \mathfrak{sp}(n)$ with bi-fundamental fermionic generators in the representation $\mathbf{m} \otimes \mathbf{n}$; the $SU(n|n)$ superalgebra has bosonic subalgebra $\mathfrak{su}(n) \oplus \mathfrak{su}(n)$ with fermionic generators in the $\mathbf{n} \otimes \mathbf{n}$. The most fancy case is the superalgebra $D(2|1; \alpha)$ which depends on a free parameter $\alpha \neq 0, -1$. $D(2|1; \alpha)$ and $D(2|1; \alpha')$ are isomorphic if and only if α and α' are in the same orbit of the order six group generated by $\alpha \rightarrow -1 - \alpha$ and $\alpha \rightarrow 1/\alpha$.

The bosonic Lie subalgebra of $D(2|1; \alpha)$ is $SU(2) \times SU(2) \times SU(2)$, while the fermionic generators form a *tri*-fundamental; so we have $9 + 8 = 17$ generators. We write the generators of the first $SU(2)$ as L_{-1}, L_0, L_{+1} and those of the other two $SU(2)$'s as $A^i, \tilde{A}^i$, respectively, so that the bosonic subalgebra reads

$$[L_m, L_n] = (m - n)L_{m+n}, \quad [A^i, A^j] = \epsilon^{ijk} A^k, \quad [\tilde{A}^i, \tilde{A}^j] = \epsilon^{ijk} \tilde{A}^k$$

$$[L_m, A^i] = [L_m, \tilde{A}^i] = [A^i, \tilde{A}^j] = 0, \quad m, n = 0, \pm 1, \quad i, j, k = 1, 2, 3.$$

We write the fermionic generators in the form $G_{+1/2}^a, G_{-1/2}^a$ with $a = 1, 2, 3, 4$. Then

$$[L_m, G_r^a] = (\tfrac{1}{2}m - r) G_{m+r}^a, \quad [A^i, G_r^a] = \eta^{ab,i} G_r^b, \quad [\tilde{A}^i, G_r^a] = \tilde{\eta}^{ab,i} G_r^b$$

($r = 1/2, -1/2, m = 0, \pm 1, i = 1, 2, 3$) where $\eta^{ab,i}$ and $\tilde{\eta}^{ab,i}$ are the self-dual and anti-self-dual 't Hooft tensors [43]. The anticommutator of two fermionic generators is

$$\{G_r^a, G_s^b\} = 2\delta^{ab} L_{r+s} + \delta_{r+s,0} (\lambda \eta^{ab,i} A^i + \tilde{\lambda} \tilde{\eta}^{ab,i} \tilde{A}^i)$$

where $\lambda, \tilde{\lambda}$ are coefficients. Rescaling the generators $G_r^a \rightarrow t G_r^a$ we make $(\lambda, \tilde{\lambda}) \rightarrow (t^2 \lambda, t^2 \tilde{\lambda})$, so the superalgebra depends (modulo isomorphism) only on the ratio $\alpha = \tilde{\lambda}/\lambda$. $D(2|1; \alpha)$ with α in the orbit of 1 is obviously isomorphic to $OSp(4|2)$.

where $\mathcal{D} = \gamma^\mu (\partial_\mu - i A_\mu)$ and A_μ is a background non-Abelian gauge connection which—for simplicity—we couple only to vector currents using a regularization which preserves the vector-like symmetries. A_μ takes value in the Lie algebra of the gauge group $G \subset O(N)$ and we see it as a matrix-valued vector field. Without loss of generality we take $G = O(N)$. We work in Euclidean signature. We set

$$W[A_\mu] \stackrel{\text{def}}{=} \log \text{Det}[i \mathcal{D}], \quad (2.570)$$

and introduce

$$J_\mu(x) \equiv \frac{\delta W}{\delta A^\mu(x)} = \langle \bar{\psi} \gamma_\mu \psi \rangle_{A_\mu} \quad (2.571)$$

where $\langle \cdots \rangle_{A_\mu}$ stands for the v.e.v. computed in the given background A_μ , and both sides take value in the Lie algebra of G . Since the vector-currents are covariantly conserved in all backgrounds

$$D^\mu J_\mu \equiv \partial^\mu J_\mu + [A^\mu, J_\mu] = 0 \quad (2.572)$$

whereas the non-Abelian axial current

$$\langle \bar{\psi} \gamma_\mu \gamma_5 \psi \rangle = \epsilon_{\mu\nu} J^\nu \quad (2.573)$$

is not conserved due to the chiral anomaly

$$\epsilon^{\mu\nu} \left(\partial_\mu J_\nu + [A_\mu, J_\nu] \right) = \text{chiral anomaly}. \quad (2.574)$$

The RHS is a *local, gauge covariant* operator in the adjoint of G , scaling as $(\text{length})^{-2}$, which reduces to $(\text{field strength})/2\pi$ (i.e. the density of first Chern class) if we restrict the background A_μ to a Cartan subalgebra. The only local operator with these properties is the non-Abelian field strength

$$\epsilon^{\mu\nu} \left(\partial_\mu J_\nu + [A_\mu, J_\nu] \right) = \frac{1}{4\pi} \epsilon^{\mu\nu} F_{\mu\nu}. \quad (2.575)$$

We have to solve Eqs. (2.572), (2.575). To do that, we use the following remark: *in (Euclidean) 2d all smooth vector bundles with connection have a complex structure compatible with that connection.*¹¹⁰ Wick rotating back to Minkowski space and using 2d light-cone coordinates $x^\pm = x_0 \pm x_1$ means that we can find two elements g, h of the gauge group G such that

$$A_+ = g^{-1} \partial_+ g \quad A_- = h^{-1} \partial_+ h. \quad (2.576)$$

Locally we may take the axial gauge, $A_- = 0$, i.e. set $h = 1$. Then we check that

$$4\pi J_+ = g^{-1} \partial_+ g \quad 4\pi J_- = -g^{-1} \partial_- g \quad (2.577)$$

is a solution to Eqs. (2.572), (2.575) in the axial gauge. Indeed,

$$4\pi \partial_+ J_- = (g^{-1} \partial_+ g)(g^{-1} \partial_- g) - g^{-1} \partial_+ \partial_- g = -4\pi A_+ J_- + 4\pi J_- A_+ + F_{+-} \quad (2.578)$$

$$4\pi \partial_- J_+ = -(g^{-1} \partial_- g)(g^{-1} \partial_+ g) + g^{-1} \partial_- \partial_+ g = F_{-+}. \quad (2.579)$$

Hence (in the axial gauge)

$$\begin{aligned} \delta W &= \int d^2 x \operatorname{tr} (J_- \delta A_+) = -\frac{1}{4\pi} \int d^2 x \operatorname{tr} \left((g^{-1} \partial_- g)(g^{-1} \partial_+ \delta g - g^{-1} \delta g g^{-1} \partial_+ g) \right) \\ &= \frac{1}{4\pi} \int d^2 x \operatorname{tr} \left(\partial_+ (g^{-1} \partial_- g) g^{-1} \delta g - (g^{-1} \partial_- g)(g^{-1} \partial_+ g) g^{-1} \delta g - (g^{-1} \partial_- g) g^{-1} \delta g (g^{-1} \partial_+ g) \right) \\ &= \frac{1}{4\pi} \int d^2 x \operatorname{tr} \left(\partial_- (g^{-1} \partial_+ g) g^{-1} \delta g \right). \end{aligned} \quad (2.580)$$

The Wess–Zumino Term On the other hand, consider the action functional ($g \in SO(N)$)

$$I[g] \equiv \frac{1}{4\lambda^2} \int d^2 x \operatorname{tr} \left(\partial_\mu g \partial^\mu g^{-1} \right) + n \Gamma \quad (2.581)$$

where Γ is the 2d Wess–Zumino term [87], a 2d renormalizable interaction which is most conveniently written as a 3d integral. To define Γ , let us continue to Euclidean signature, and take spacetime to be a large 2-sphere S . Since $\pi_2(SO(N)) = 0$, a mapping g from S into the group $SO(N)$ may be extended to a map g from the solid ball B with $\partial B = S$ to $SO(N)$. Let $y^i, i = 1, 2, 3$ be the coordinates of B . Then the Wess–Zumino functional is

$$\Gamma = \frac{1}{24\pi} \int_B d^3 y \epsilon^{ijk} \operatorname{tr} \left((g^{-1} \partial_i g)(g^{-1} \partial_j g)(g^{-1} \partial_k g) \right), \quad (2.582)$$

¹¹⁰ Recall that a smooth vector bundle on a complex manifold is *holomorphic* iff it admits a connection A such that its curvature $F = dA + A^2$ has vanishing $(2, 0)$ component. The last condition is automatic in 1 complex dimension (2 real dimensions).

i.e. $\Gamma = C \int \text{tr}(g^{-1} dg)^3$ with C a suitable normalization constant. So defined, the Wess–Zumino functional is well-defined only up to an additive constant: the ambiguity in Γ arises because there are topologically inequivalent ways of extending g to a map from B to $SO(N)$. Giving two such extensions $B_1 \rightarrow SO(N)$, $B_2 \rightarrow SO(N)$ which agree on the boundary, we may glue them along the boundary (inverting the orientation of one of the two) to get a map from S^3 to $SO(N)$. Then

$$\Delta\Gamma = C \int_{B_2} \text{tr}(g^{-1} dg)^3 - C \int_{B_1} \text{tr}(g^{-1} dg)^3 = C \int_{S^3} \text{tr}(g^{-1} dg)^3 \equiv C' \int_{S^3} g^* \varrho_3 \quad (2.583)$$

where ϱ_3 is a 3-form representing the generator of $H^3(\text{Spin}(N), \mathbb{Z})$. The normalization in Eq. (2.582) is chosen so that the ambiguity $\Delta\Gamma$ is an integer multiple of 2π if g is a matrix in the fundamental representation of $SO(N)$. Γ is a good 2d Lagrangian term since it may be written as an integral of a density over spacetime; indeed, locally on $SO(N)$ $\varrho_3 = d\lambda_2$ and $\Gamma = \int_{\text{spacetime}} g^* \lambda_2$. However, since λ_2 is not globally defined, under an $SO(N) \times SO(N)$ transformation it will change by an exact term $\lambda_2 \rightarrow \lambda_2 + d\xi$, so that it is $SO(N) \times SO(N)$ symmetric but not *manifestly* so. If $\exp(i\Gamma)$ has to be univalued, the overall coefficient of Γ in Eq. (2.581) should be an integer. The variation of Γ is a local functional in the 2d spacetime. Indeed,

$$\begin{aligned} \delta \int_B \text{tr}(g^{-1} dg)^3 &= \int_B \text{tr} \left\{ (g^{-1} dg)^2 \delta(g^{-1} dg) \right\} = \int_B \text{tr} \left\{ (g^{-1} dg)^2 \left[- (g^{-1} \delta g)(g^{-1} dg) + g^{-1} \delta dg \right] \right\} = \\ &= \int_B \text{tr} \left\{ (g^{-1} dg)^3 (-g^{-1} \delta g) \right\} + \int_B \text{tr} \left\{ (g^{-1} dg)^2 (g^{-1} d\delta g) \right\} = \int_B \text{tr} \left\{ (g^{-1} dg)^2 d(g^{-1} \delta g) \right\} = \\ &= - \int_B \text{tr} \left\{ d(g^{-1} dg) d(g^{-1} \delta g) \right\} = \int_B d \text{tr} \left\{ d(g^{-1} dg) (g^{-1} \delta g) \right\} = \int_S \text{tr} \left\{ d(g^{-1} dg) (g^{-1} \delta g) \right\}. \end{aligned} \quad (2.584)$$

Reinserting the normalizations, this becomes

$$\delta\Gamma = -\frac{1}{8\pi} \int d^2x \text{tr} \left[(g^{-1} \delta g) \epsilon^{\mu\nu} \partial_\mu (g^{-1} \partial_\nu g) \right]. \quad (2.585)$$

Returning to Eq. (2.581), we have

$$\begin{aligned} \delta I &= \int d^2x \text{tr} \left\{ (g^{-1} \delta g) \left(\frac{1}{2\lambda^2} + \frac{n}{8\pi} \right) \partial_- (g^{-1} \partial_+ g) \right\} + \\ &+ \int d^2x \text{tr} \left\{ (g^{-1} \delta g) \left(\frac{1}{2\lambda^2} - \frac{n}{8\pi} \right) \partial_+ (g^{-1} \partial_- g) \right\}. \end{aligned} \quad (2.586)$$

The Fermionic Determinant Comparing with Eq. (2.580) we see that in the axial gauge $A_- = 0$ and parameterization $A_+ = g^{-1} \partial_+ g$, the determinant of the fermionic integral is given by

$$\log \text{Det}[\gamma^\mu (\partial_\mu + A_\mu)] \equiv W[A_\mu] = I[g] \Big|_{\substack{n=1 \\ \lambda^2=4\pi}}. \quad (2.587)$$

The expression in a generic gauge may be easily obtained using the fact that our regulation preserves the vector gauge invariance

$$g(x) \rightarrow g(x)U(x), \quad h(x) \rightarrow h(x)U(x), \quad (2.588)$$

so the functional determinant depends only on the gauge-invariant combination gh^{-1} . Then

$$\log \text{Det}[\gamma^\mu (\partial_\mu + A_\mu)] \equiv W[A_\mu] = I[gh^{-1}] \Big|_{\substack{n=1 \\ \lambda^2=4\pi}}. \quad (2.589)$$

However, the left (resp. right) fermions couple only to A_- (resp. A_+) and hence—up to local counterterms needed to enforce gauge invariance under *vector* gauge transformations— $\log \text{Det } \mathcal{D}$ should be of the form $F[g] + F[h^{-1}]$, where $F[\cdot]$ is $I[\cdot]$ at $n = 1$ and $\lambda^2 = 4\pi$. Indeed, one finds

$$W[A_\mu] = W[A_+] + W[A_-] + \frac{1}{4\pi} \int d^2x \operatorname{tr}(A_+ A_-) \quad (2.590)$$

or

$$I[g] + I[h^{-1}] + \frac{1}{4\pi} \int d^2x \operatorname{tr}[(g^{-1} \partial_+ g)(h^{-1} \partial_- h)] = I[gh^{-1}] \quad (2.591)$$

which is the *Polyakov–Wiegmann identity* [86].

The Wess–Zumino Functional Of course, the functional $W[A_\mu]$ is not invariant under axial gauge transformations

$$g(x) \rightarrow g(x)U(x), \quad h(x) \rightarrow h(x)U(x)^{-1} \quad (2.592)$$

because of the anomaly. Writing A_μ^U for the gauge connection obtained by the replacement (2.592), we define the *Wess–Zumino functional (or action)* $WZ[U, A_\mu]$ as [87]

$$WZ[U^2, A_\mu] \equiv W[A_\mu^U] - W[A_\mu] \quad (2.593)$$

which is a global version of the axial anomaly. Using the Polyakov–Wiegmann identity (2.591)

$$\begin{aligned} WZ[U; g, h] &= I[gUh^{-1}] - I[gh^{-1}] = I[gU] + I[h^{-1}] + \\ &+ \frac{1}{4\pi} \int d^2x \operatorname{tr}[(U^{-1}g^{-1}\partial_+(gU))(h^{-1}\partial_-h)] - I[g] - I[h^{-1}] - \frac{1}{4\pi} \int d^2x \operatorname{tr}[(g^{-1}\partial_+g)(h^{-1}\partial_-h)] \\ &= I[U] + \frac{1}{4\pi} \int d^2x \operatorname{tr}[(g^{-1}\partial_+g)(U\partial_-U^{-1})] + \frac{1}{4\pi} \int d^2x \operatorname{tr}[(U^{-1}g^{-1}\partial_+(gU) - g^{-1}\partial_+g)(h^{-1}\partial_-h)] \\ &= I[U] + \frac{1}{4\pi} \int d^2x \operatorname{tr}\{A_+ U \partial_- U^{-1} + U^{-1} \partial_+ U A_- + U^{-1} A_+ U A_- - A_+ A_-\}. \end{aligned} \quad (2.594)$$

Bosonization Identities The bosonization identity on the generating functional for the current correlations reads

$$\begin{aligned} \int [d\psi \, d\bar{\psi}] \exp\left\{-\int d^2x \, i \, \bar{\psi} \mathcal{D} \psi\right\} &\equiv \\ &\equiv \text{Det}[\mathcal{D}] \equiv \exp\{I[gh^{-1}]\} = \int [dU] \exp\left\{-WZ[U, A_\mu]\right\} \end{aligned} \quad (2.595)$$

which relates the system of free fermions to a bosonic sigma-model with target space the group manifold G in a background gauge field A_μ . The proof is simple

$$1 = \int [dU] e^{-I[U]} = \int [dU] e^{-I[gh^{-1}]} = e^{-I[gh^{-1}]} \int [dU] \exp\left\{-WZ[U, A_\mu]\right\}, \quad (2.596)$$

where the first equality is a choice of normalization of the measure.

Hence the correlation functions of left (or right) currents in the free fermionic theory

$$\begin{aligned} \langle J_+(x_1) \dots J_+(x_n) \rangle &= \frac{\delta^n e^{W[A_\mu]}}{\delta A_-(x_1) \dots \delta A_-(x_n)} \Big|_{A_\pm=0} = \\ &= \frac{(-1)^n}{(4\pi)^n} \langle U^{-1} \partial_+ U(x_1) \dots U^{-1} \partial_+ U(x_n) \rangle_{\text{bosonic}} \end{aligned} \quad (2.597)$$

so that we get the operator identifications (Euclidean signature)

$$j_+ \rightsquigarrow -\frac{1}{4\pi} U^{-1} \partial_+ U \quad j_- \rightsquigarrow -\frac{1}{4\pi} U \partial_- U^{-1} = \frac{1}{4\pi} (\partial_- U) U^{-1} \quad (2.598)$$

which are the Witten bosonization rules [85] (up to different conventions—we use the ones in [88]). Notice also that we may have a discrepancy at coinciding arguments, meaning a different definition of T -ordering (which does not change the physical content of the theory).

The actual meaning of these identities is that both the free massless fermions and the Wess–Zumino bosonic theory at level $n = 1$ are conformal theories which lead to unitary representations of the level 1 affine $O(N)$ algebra, which is essentially unique, and hence have isomorphic Hilbert spaces and the same energy–momentum tensor (that is, the same dynamics) which, in the fermionic setup, is just the Sugawara one

$$T(z) = \frac{1}{k + h(G)^\vee} : J^a(z) J^a(z). \quad (2.599)$$

To complete the bosonization dictionary we need a rule for the scalar Fermi bilinears. Reinstating the gauge indices, the straightforward generalization of the Abelian case is

$$i \psi_-^k \psi_{j+}(x) = \Lambda U(x)^k_j, \quad (2.600)$$

where Λ is an RG-prescription-dependent scale. That this identification is correct in general may be seen in many ways [85, 88].

The Wess–Zumino–Witten (WZW) CFT

Let G be a compact Lie group. The action (2.581) with $\lambda^2 = 4\pi$ and $g \in G$ is called the Wess–Zumino–Witten model (WZW) with group G and level n (in the general case by tr we mean the trace normalized so that Γ is well defined mod 1). The WZW yields an explicit Lagrangian formulation of the pure 2d current algebra CFT associated with the level- n Kač–Moody symmetry G_n in terms of bosonic fields. Its energy–momentum tensor coincides with the Sugawara one.

Appendix 2: Valued Graphs, Affine Lie Algebras, McKay Correspondence, and All That

A *Cartan matrix of rank n* C_{ij} (in the sense of Kač [52]) is an integral $n \times n$ matrix with diagonal entries equal to 2 and $C_{ij} \leq 0$ for $i \neq j$ while $C_{ij} \neq 0$ implies $C_{ji} \neq 0$. A Cartan matrix is *simply laced* iff it is symmetric. Given a $n \times n$ Cartan matrix C , one constructs a Lie algebra $\mathfrak{g}_C$ (called a *Kač–Moody algebra*) by the usual construction [52, 54]: $\mathfrak{g}_C$ is the Lie algebra over $3n$ generators e_i, f_i, h_i ($i = 1, \dots, n$) with relations

$$[e_i, f_j] = \delta_{ij} h_i, \quad [h_i, e_j] = C_{ij} e_j, \quad [h_i, f_j] = -C_{ij} f_j, \quad (2.601)$$

$$[h_i, h_j] = 0, \quad (\text{ad } e_i)^{1-C_{ij}} e_j = (\text{ad } f_i)^{1-C_{ij}} f_j \quad \text{for } i \neq j. \quad (2.602)$$

There are three possibilities [52]:

- F** C is positive-definite: $\mathfrak{g}_C$ is a *finite-dimensional* Lie algebra;
- N** C is semi-positive-definite with one zero eigenvalue: $\mathfrak{g}_C$ is an *affine* Lie algebra, i.e. a 2d current algebra for some Lie group G_0 ;
- I** C is indefinite and non-singular.

We consider only simply laced algebras, the non-simply laced ones can be obtained by a “folding” procedure which generalizes the one in Fig. 2.2 for the case **F**.

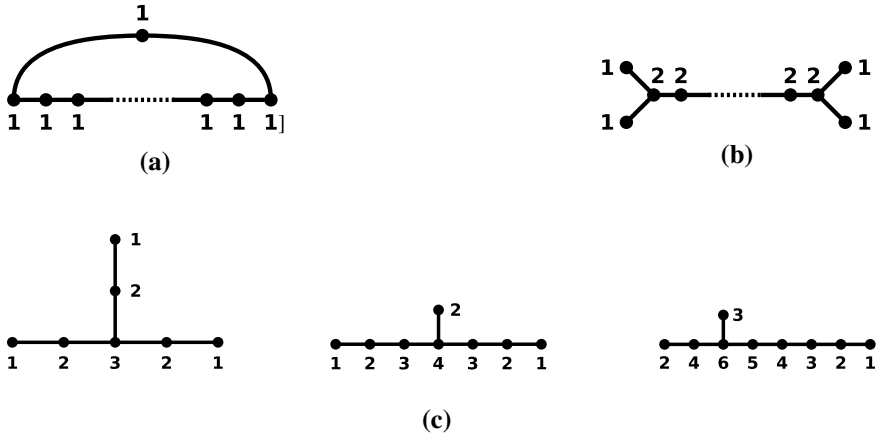


Fig. 2.3 Affine Dynkin graphs (with Coxeter labels) of exceptional type $\tilde{E}_r$ for $r = 6, 7, 8$

An unoriented *graph* Γ is specified by its set of *nodes* (vertices), labeled by elements of the set $\mathbf{N} \equiv \{1, 2, \dots, n\}$, together with a symmetric *incidence function*

$$I: \mathbf{N} \times \mathbf{N} \longrightarrow \mathbb{Z}_{\geq 0}, \quad (i, j) \mapsto I_{ij} \in \mathbb{Z}_{\geq 0}, \quad (2.603)$$

where $I_{ij} = I_{ji}$ is the number of *links* (edges) in Γ which connect the nodes i and j . We see I_{ij} as a $n \times n$ symmetric matrix. The diagonal entry I_{ii} is the number of *loops* based at i ($\equiv$ edges starting and ending at i). We consider only *loop-free* graphs, i.e. with $I_{ii} = 0$ for all i . The graph is *simply laced* iff $I_{ij} \in \{0, 1\}$. The Cartan matrix of Γ is

$$C_{ij} = 2\delta_{ij} - I_{ij}. \quad (2.604)$$

Equation (2.604) sets a correspondence between simply laced Cartan matrices C (hence simply laced Kač–Moody Lie algebras) and simply laced graphs Γ_C . Γ_C is the *Dynkin graph* of the Lie algebra $\mathfrak{g}_C$. We may assume Γ_C to be connected $\Leftrightarrow$ the Lie algebra $\mathfrak{g}_C$ to be *simple*.

A simply laced graph Γ corresponds to an *affine* Lie algebra iff there exists a function¹¹¹

$$v: \mathbf{N} \rightarrow \mathbb{N} \quad \text{such that} \quad 2v_i = I_{ij}v_j \text{ and } \gcd(v_i) = 1. \quad (2.605)$$

The positive integer v_i attached to the i th node is the *Coxeter label* of the node.

Claim 2.5 The connected graphs with a function v with the properties (2.605) are the simply laced affine graphs of types $\tilde{A}_n, \tilde{D}_n, \tilde{E}_r$ graphs ($n \in \mathbb{N}, r = 6, 7, 8$) in Fig. 2.3a–c. The function $v: \mathbf{N} \rightarrow \mathbb{N}$ associates with each node the integer drawn near it. The pair (Γ, v) is a *valued graph*.

A node i of an affine graph is an *extension node* iff $v_i = 1$. From the figures we check that:

¹¹¹ **Proof.** By condition **N**, C_{ij} has a zero eigenvector v which is rational, so a multiple of it has integral entries with $\gcd(v_i) = 1$. It remains to show that these entries are positive. A graph is connected iff there is an integer m such that the m th power of its incidence matrix I^m has strictly positive entries. By the Perron–Frobenius theorem [89] the entries of the eigenvector v_{pf} of I^m associated with its largest eigenvalue λ_{pf} are strictly positive. By condition **N** the spectrum of I_{ij} is bounded above by 2, therefore $v \equiv v_{\text{pf}}$ is the eigenvector associated with the largest eigenvalue $\lambda_{\text{pf}} = 2^m$ of I^m , hence it has strictly positive entries.

Claim 2.6 (1) The automorphism group of an affine diagram Γ acts transitively on the extension nodes. (2) Deleting one extension node of the affine graph $\tilde{\mathfrak{g}}$ we get the Dynkin graph $\Gamma_{\mathfrak{g}}$ of the corresponding finite-dimensional Lie algebra $\mathfrak{g}$. (3) The number of extension nodes is the order of the center of the simply connected Lie group G with Lie algebra $\mathfrak{g}$. (4) The Coxeter number ($\equiv$ the dual Coxeter number for a simply laced Lie algebra) of $\mathfrak{g}$ is equal to $\sum_i v_i$ for the corresponding affine graph $\tilde{\mathfrak{g}}$. (5) The largest root of a simply laced finite-dimensional Lie algebra $\mathfrak{g}$ is¹¹²

$$\theta = \sum_{i \in \Gamma_{\mathfrak{g}}} v_i \alpha_i. \quad (2.606)$$

Roots and All That Let C_{ij} be the Cartan matrix of a simply laced graph Γ with n nodes.

Definition 2.2 with connected support in Γ such that $w^t C w \leq 2$. A root is *real* if the inequality is saturated. Otherwise the root is *imaginary*.

In case **F**, C is positive-definite, so there are only finitely many roots all real. In the affine case **N**, the imaginary roots correspond to the radical of the Tits quadratic form $q(w) \equiv \frac{1}{2} w^t C w$,

$$\text{rad } q = \{w \in \mathbb{Z}^n : q(w) = 0\}, \quad (2.607)$$

which is the rank-1 lattice $\mathbb{Z}\delta$, where δ is the fundamental imaginary root $\delta \equiv \sum_i v_i \alpha_i$. The real roots have the form

$$\alpha + n\delta, \quad \alpha \in \Delta(\mathfrak{g}) \text{ (a root of } \Gamma_{\mathfrak{g}}), \quad n \in \mathbb{Z}. \quad (2.608)$$

Reconstructing the Affine Lie Algebra ($\equiv$ Current Algebra)

We now reconstruct the affine Lie algebra $\tilde{\mathfrak{g}}$ out of the affine graph Γ (equivalently, out of its affine Cartan matrix C_{ij}). The rank of the affine Lie algebra is $r + 1$, where r is the rank of the corresponding finite-dimensional Lie algebra $\mathfrak{g}$ (the extended graph $\tilde{\Gamma}$ has one more node than the Dynkin graph of $\mathfrak{g}$ which is an extension node). The Cartan algebra $\tilde{\mathfrak{h}}$ of $\tilde{\mathfrak{g}}$ is then

$$\tilde{\mathfrak{h}} = \mathfrak{h} \oplus \mathbb{C} \cdot K \quad (2.609)$$

with K central (a number in an irreducible representation). $\tilde{\mathfrak{g}}$ decomposes into *root eigenspaces*:

$$\tilde{\mathfrak{g}} = \tilde{\mathfrak{h}} \oplus \left(\bigoplus_{\alpha: \text{root}} \tilde{\mathfrak{g}}^{\alpha} \right) \quad \dim_{\mathbb{C}} \tilde{\mathfrak{g}}^{\alpha} = \begin{cases} 1 & \alpha \text{ real} \\ r & \alpha \text{ imaginary.} \end{cases} \quad (2.610)$$

We write $H^i(z)$ and $E^{\alpha}(z)$ for the holomorphic currents of a CFT with left-moving symmetry G , corresponding, respectively, to the elements of the Cartan algebra and the α -root space of the finite-dimensional Lie algebra $\mathfrak{g} \equiv \mathfrak{Lie}(G)$. We have

$$H^i(z) = \sum_{n \in \mathbb{Z}} \frac{H_n^i}{z^{n+1}}, \quad E^{\alpha}(z) = \sum_{n \in \mathbb{Z}} \frac{E_n^{\alpha}}{z^{n+1}}. \quad (2.611)$$

Then, for α a root of $\mathfrak{g}$

$$\tilde{\mathfrak{g}}^{n\delta+\alpha} = \mathbb{C} \cdot E_n^{\alpha}, \quad \tilde{\mathfrak{g}}^{n\delta} = \text{span of } (H_n^1, \dots, H_n^r) \quad (2.612)$$

while the $r + 1$ simple roots are

$$\{\alpha_0 = K - \theta, \alpha_1, \alpha_2, \dots, \alpha_r\} \quad (2.613)$$

with α_i ($i \geq 1$) the simple roots of $\mathfrak{g}$ and $\theta = \sum_{i \geq 1} m_i \alpha_i$ the maximal root ($m_i \equiv v_i$ for $i \geq 1$). The simple co-roots are

¹¹² As always, α_i stands for the simple root associated with the i th node of Γ .

$$\{K - \theta^\vee, \alpha_1^\vee, \dots, \alpha_r^\vee\}, \quad (2.614)$$

so the affine Cartan matrix is

$$\tilde{C} = \left(\begin{array}{c|c} 2 & -m_k C_{kj} \\ \hline -C_{ik} m_k & C_{ij} \end{array} \right), \quad i, j = 1, \dots, r, \quad (2.615)$$

where C_{ij} is the Cartan matrix of the Lie algebra $\mathfrak{g}$. From the Lie algebra relations we get

$$[E_n^\alpha, E_n^{-\alpha}] = (n\delta + \alpha)^\vee = nK + H_0^\alpha, \quad (2.616)$$

so we see that the affine algebra is the central extension of the loop algebra of G , hence (by uniqueness) the 2d current algebra with group G .

The McKay Correspondence

The characterization of the simply laced affine Dynkin graphs as the graphs with a positive function v such that $2v_i = I_{ij}v_j$ implies that the classification of the simply laced affine Lie algebras,

$$\tilde{A}_r, \tilde{D}_r, \tilde{E}_6, \tilde{E}_7, \tilde{E}_8, \quad (2.617)$$

enters every time the classification at hand may be reduced to listing graphs with such a function. There are several dozens of fundamental classification problems, in pure mathematics as well as in physics, which follow this “ADE pattern” [90–92]. Examples are the Kodaira classification of singular fibers in elliptic fibrations [93–95], the classification of 2d CFTs with $c < 1$ [96], of 2d $\mathcal{N} = 2$ SCFTs with $c < 3$ [97, 98], and the classification of finite subgroups of $SU(2)$ [99]. The following fundamental result was essentially known to Plato in the fourth century B.C.

Theorem 2.3 (McKay [57, 99]) *The finite subgroups of $SU(2)$ are in one-to-one correspondence with the simply laced affine Lie algebras $\tilde{A}_n, \tilde{D}_n, \tilde{E}_r$ with $n \in \mathbb{N}$ and $r = 6, 7, 8$.*

Proof Saying that a finite group G is a subgroup of $SU(2) \subset GL(2, \mathbb{C})$ is equivalent to saying that it has a faithful quaternionic¹¹³ Q representation of dimension 2. If Q is reducible, $Q = \psi \oplus \psi^\vee$, with ψ one-dimensional, and G is a finite subgroup of $U(1)$, hence $G \simeq \mathbb{Z}_n$ for some n . Otherwise G is non-Abelian and $-1 \in G$.¹¹⁴ In either case, the center $Z(G)$ contains a non-trivial element ζ .

Let $\{R_i\}$ ($i = 0, 1, \dots, m$) be the set of (pair-wise non-isomorphic) irreducible representations R_i of G , with R_0 the trivial representation. We write

$$Q \otimes R_i = \bigoplus_{j=0}^m M_{ij} R_j \quad (2.618)$$

¹¹³ A linear representation R is called *quaternionic* (a.k.a. symplectic or, in the physical literature, *pseudo-real*) iff its character $\chi(g)$ takes only real values (i.e. the representation R is isomorphic to its dual $R^\vee$) and the invariant $R \otimes R \rightarrow \mathbb{C}$ is *antisymmetric*. An irreducible representation R is quaternionic if and only if its Frobenius–Schur indicator [28] satisfies $\frac{1}{|G|} \sum_{g \in G} \chi(g^2) = -1$.

¹¹⁴ **Proof.** If G is not solvable, G should contain an involution (hence -1) by the Feit–Thompson theorem [100]. Otherwise, G is of the form $A \rtimes H$ with A an Abelian normal subgroup. Hence $A \simeq \mathbb{Z}_n$ and the action of H on $\mathbb{Z}_n$ factors through $\text{Aut}(\mathbb{Z}_n) \simeq \mathbb{Z}_2$, so H has even order. Therefore $|G| = n|H|$ is even, and thus G contains an involution by Sylow’s theorem [101].

where the non-negative integer M_{ij} is the multiplicity of the representation R_j in the representation $Q \otimes R_i$. We write χ_i for the character¹¹⁵ of R_i and χ for the character of Q . Equation (2.618) translates in the relation between characters¹¹⁶

$$\chi(g) \chi_i(g) = M_{ij} \chi_j(g) \quad g \in G, \quad (2.619)$$

that is, for fixed $g \in G$ the vector $(\chi_0(g), \chi_1(g), \dots, \chi_m(g))^t$ is an eigenvector of the integral matrix M_{ij} , associated with the eigenvalue $\chi(g)$ which is *real* (since Q is symplectic) with

$$|\chi(g)| \leq \dim Q \equiv 2 \quad \text{for all } g \in G. \quad (2.620)$$

The matrix M_{ij} is symmetric with zeros on the diagonal. Indeed, by orthogonality of the characters

$$\begin{aligned} \frac{1}{|G|} \sum_{g \in G} \chi_i(g)^* \chi(g) \chi_j(g) &= M_{ij} \equiv M_{ij}^* = \\ &= \frac{1}{|G|} \sum_{g \in G} \chi_i(g) \chi(g)^* \chi_j(g)^* = \frac{1}{|G|} \sum_{g \in G} \chi_i(g) \chi(g) \chi_j(g)^* = M_{ji} \end{aligned} \quad (2.621)$$

since $\chi(g)$ is real. Now let $\zeta \in Z(G)$, $\zeta \neq 1$. $\chi_j(\zeta) = \zeta^{n_j} \dim R_j$ for some $n_j \in \mathbb{Z}$, and hence

$$M_{ij} = 0 \quad \text{unless } n_j - n_i = 1 \bmod |Z(G)| \quad (2.622)$$

which implies $M_{ii} = 0$. We conclude that the non-negative matrix $M_{ij} \equiv I_{ij}$ is the incidence matrix of an unoriented, loop-free graph Γ . Now specialize Eq. (2.619) to $g = 1$; we get

$$2 \dim R_i = M_{ij} \dim R_j \quad (2.623)$$

so that $v_i = \dim R_i$ is a function on Γ with the properties (2.605) and then, by **Claim 2.5** Γ is an affine Dynkin graph of one of the types $\tilde{A}\tilde{D}\tilde{E}$. $\square$

Relation to Canonical Singularities

The McKay correspondence is strictly related to the theory of canonical surface singularities a.k.a. du Val singularities, minimal singularities, or *ADE* singularities [103–106]. Let $G \subset SU(2)$ be a non-trivial finite subgroup and consider the quotient space

$$X_G \stackrel{\text{def}}{=} \mathbb{C}^2 / G \quad (2.624)$$

which is singular at the origin $0 \in \mathbb{C}^2$. The symplectic form $dz_1 \wedge dz_2$ is G -invariant hence well defined on the quotient space $\mathbb{C}^2 / G$. This suggests that the singularity admits a *crepant resolution*, i.e. a holomorphic map $h: Y_G \rightarrow \mathbb{C}^2 / G$ from a smooth complex surface Y_G which is an isomorphism away from the origin and such that the canonical bundle K_{Y_G} is the pull-back of the one on the quotient (hence trivial).

One way to study the geometry of the singular quotient $\mathbb{C}^2 / G$ is to study the ring of polynomial function in $\mathbb{C}^2$ which are G -invariant $\mathcal{R}^G \equiv \mathbb{C}[z_1, z_2]^G$. A celebrated theorem by Klein [107] states that, for $G \subset SU(2)$, $\mathcal{R}^G$ is a graded $\mathbb{C}$ -algebra generated by three invariants x_1, x_2 , and x_3 related by a single polynomial relation. For instance for the cyclic group $\mathbb{Z}_{n+1} \subset SU(2)$, we have $x_1 = z_1^{n+1}$, $x_2 = -z_2^{n+1}$, $x_3 = z_1 z_2$ and we have the relation

$$x_1 x_2 + x_3^{n+1} = 0, \quad (2.625)$$

¹¹⁵ For background about linear representations of finite groups and their characters, see, for example, [102].

¹¹⁶ Here we exploit the fact that the character χ gives an isomorphism between the representation ring and the ring of central functions on the group G [102].

Table 2.2 Canonical (a.k.a. du Val) surface singularities and Lie algebras

Finite group	$\mathbb{Z}_{n+1}$	Binary dihedral	Binary tetrahedral	Binary octahedral	Binary icosahedral
Lie algebra	A_n	D_n	E_6	E_7	E_8
Relation	$x_1 x_2 + x_3^{n+1}$	$x_1^2 + x_2^2 x_3 + x_3^{n-1}$	$x_1^2 + x_2^2 + x_3^4$	$x_1^2 + x_2^3 + x_3^3$	$x_1^2 + x_2^3 + x_3^5$

while for the binary dihedral group with $4(n-2)$ (with, say, n odd) $x_1 = (z_1^{2n-3} z_2 - z_2^{2n-3} z_1)/2$, $x_2 = i(z_1^{2n-4} + z_2^{2n-4})/2$, $x_3 = (z_2 z_1)^2$ with the relation

$$x_1^2 + x_2^2 x_3 + x_3^{n-1} = 0. \quad (2.626)$$

The polynomial relation between the 3 basic invariants is the equation of the singular surface $\mathbb{C}^2/G$ as a hypersurface in $\mathbb{C}^3$. These surfaces are singular, and the polynomials $W(x_1, x_2, x_3)$ which describe them are called the minimal (or canonical, or du Val) singularities [103–106]. It is convenient to label them with the finite-dimensional simply laced Lie algebra $\mathfrak{g}$ which correspond to them via the McKay correspondence; see Table 2.2. We are interested in the geometry of the resolved hypersurface Y_G in particular its topology. For a singular surface with an isolated singularity, the *Milnor theorem* [108] states that the resolution Y_G is diffeomorphic to the deformed smooth hypersurface $W(x_1, x_2, x_3) = \epsilon$ (locally around the origin). In particular the $h^{-1}(0) \subset Y_G$ is a bouquet of r 2-spheres (i.e. rational curves $\mathbb{P}^1$ [106]), where r is the rank of the Lie algebra $\mathfrak{g}$. The configuration of rational curves is dual to the the Dynkin graph of $\mathfrak{g}$ meaning that to each node of the graph there corresponds a sphere, and the two spheres associated with the i th and j th nodes intersect transversely in a number of points equal to the number of edges between the corresponding nodes (i.e. in $|C_{ij}|$ points). See [104] for illuminating examples.

References

1. Belavin, A.A., Polyakov, A.M., Zamolodchikov, A.B.: Infinite conformal symmetry in two-dimensional quantum field theory. Nucl. Phys. B **241**, 333–380 (1984)
2. Friedan, D., Martinec, E., Shenker, S.: Conformal invariance, supersymmetry, and string theory. Nucl. Phys. B **271**, 93 (1986)
3. Di Francesco, P., Mathieu, P., S  n  chal, D.: Conformal Field Theory. Springer, Berlin (1997)
4. Itzykson, C., Saleur, H., Zuber, J.-B. (eds.): Conformal Invariance and Applications to Statistical Mechanics. World Scientific (1988)
5. Itzykson, C., Drouffe, L.-M.: Statistical Field Theory, vol. 2. Cambridge University Press (1989)
6. Henkel, M.: Conformal Invariance and Critical Phenomena. Springer, Berlin (1999)
7. Blumenhagen, R., Plauschinn, E.: Introduction to Conformal Field Theory: With Applications to String Theory. Lecture Notes in Physics, vol. 779. Springer, Berlin (2009)
8. Ginsparg, P.: Applied conformal field theory. In: Brezin, E., Zinn-Justin, J. (eds.) Fields, Strings and Critical Phenomena. Les Houches 1988, pp. 1–168. North-Holland (1992). [arXiv:hep-th/9108028](https://arxiv.org/abs/hep-th/9108028)
9. Gaberdiel, M.: An introduction to conformal field theory. Rep. Prog. Phys. **63**, 607 (2000)
10. Cardy, J.: Conformal Field Theory and Statistical Mechanics. [arXiv:0807.3472](https://arxiv.org/abs/0807.3472)
11. Kobayashi, S.: Transformation Groups in Differential Geometry. Classics in Mathematics. Springer, Berlin (1995)

12. Postnikov, M.M.: *Geometry VI: Riemannian Geometry*. Encyclopaedia Mathematica Sciences, vol. 91. Springer, Berlin (2001)
13. Anspinnwall, P.S., Bridgeland, T., Craw, A., Douglas, M.R., Gross, M., Kapustin, A., Moore, G.W., Segal, G., Szendrői, B., Wilson, P.M.H.: *Dirichlet Branes and Mirror Symmetry*. Clay Mathematical Monographs, vol. 4. AMS (2009)
14. Yano, K.: *The Theory of Lie Derivatives and its Applications*. Nabu Press (2011)
15. Hörmander, L.: *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis*, 2nd ed. Springer, Berlin (1990)
16. Bott, R., Tu, L.W.: *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics, vol. 82. Springer, Berlin (1982)
17. Schoen, R., Yau, S.-T.: *Lectures on Differential Geometry*. In: *Conference Proceedings and Lecture Notes in Geometry and Topology. I*. International Press, Cambridge, MA (1994)
18. Aubin, T.: *Some Nonlinear Problems in Riemannian Geometry*. Springer Monographs in Mathematics. Springer, Berlin (1998)
19. Grauert, H., Remmert, R.: *Theory of Stein Spaces*. Classics in Mathematics. Springer, Berlin (1979)
20. Griffiths, P., Harris, J.: *Principles of Algebraic Geometry*. Wiley (1978)
21. Friedan, D.: In: Zuber, J.B., Stora, R. (eds.) *Les Houches 1982. Recent advances in Field Theory and Statistical Physics*. North-Holland (1984)
22. Rychkov, S.: *EPFL Lectures on Conformal Field Theory in $D \geq 3$ Dimensions* (2016). [arXiv:1601.05000](https://arxiv.org/abs/1601.05000)
23. El-Showk, S.: *Solving Conformal Theories with the Bootstrap: Overview and Recent Results*, Lectures at the 9th Asian Winter School on Strings, Particles and Cosmology, January 2015, Busan, South Korea. <http://home.kias.re.kr/MKG/h/AWSSPC2015/~pageNo=1006>
24. Qualls, J.D.: *Lectures on Conformal Field Theory* (2015). [arXiv:1511.04074](https://arxiv.org/abs/1511.04074)
25. Simmons-Duffin, D.: *TASI Lectures on the Conformal Bootstrap*. [arXiv:1602.07982](https://arxiv.org/abs/1602.07982)
26. Carathéodory, C.: *Theory of Functions of a Complex Variable*, vol. 2. Chelsea Publishing Company (1983)
27. Streater, R.F., Wightman, A.S.: In: Benjamin, W.A. (ed.) *PCT, Spin and Statistics, and All That* (1964)
28. Fulton, W., Harris, J.: *Representation Theory: A First Course*. Graduate Texts in Mathematics, vol. 129. Springer, Berlin (1991)
29. Bourbaki, N.: *Éléments de mathématique. Groupes et algèbres de Lie. Chapitre 1: Algèbres de Lie*. Springer, Berlin (2007)
30. On-Line Encyclopaedia of Integer Sequence. <https://oeis.org>
31. Andrews, G.E., Eriksson, K.: *Integer Partitions*, 2nd revised edn. Cambridge University Press (2004)
32. Andrews, G.E.: *The Theory of Partitions*. Cambridge University Press (2010)
33. Fuchs, J., Runkel, I., Schweigert, C.: Twenty-five years of two-dimensional rational conformal field theory. *J. Math. Phys.* **51**, 015210 (2010). [arXiv:0910.3145](https://arxiv.org/abs/0910.3145)
34. Cardy, J.: Operator content of two-dimensional conformal invariant theory. *Nucl. Phys. B* **270**, 186 (1986)
35. Coleman, S.R.: There are no Goldstone bosons in two-dimensions. *Commun. Math. Phys.* **31**, 259–264 (1973)
36. Mermin, N.D., Wagner, H.: Absence of ferromagnetism or antiferromagnetism in one- or two-dimensional isotropic Heisenberg models. *Phys. Rev. Lett.* **17**, 1133 (1966)
37. Weinberg, S.: *The Quantum Theory of Fields*, vols. 1, 2, 3. Cambridge University Press (1995)
38. Friedan, D.: Notes on string theory and two-dimensional conformal field theory. In: Santa Barbara, C.A., Green, M., Gross, D. (eds.) *Workshop on Unified String Theories*, vol. 1986, pp. 162–213. World Scientific, Singapore (1985)
39. Pauli, W.: The connection between spin and statistics. *Phys. Rev.* **58**, 716–722 (1940)
40. Coleman, S.: Quantum sine-Gordon equation as the massive Thirring model. *Phys. Rev. D* **11**, 2088 (1975)

41. Mandelstam, S.: Soliton operators for the quantized sine-Gordon equation. *Phys. Rev. D* **11**, 3026 (1975)
42. Alvarez-Gaume, L., Bost, J.B., Moore, G.W., Nelson, P.C., Vafa, C.: Bosonization on higher genus Riemann surfaces. *Commun. Math. Phys.* **112**, 503 (1987)
43. Hooft, G.: Computation of the quantum effects due to a four-dimensional pseudoparticle. *Phys. Rev.* **D14**, 3432–3450 (1976)
44. Coleman, S.R.: The uses of instantons. *Subnucl. Ser.* **15**, 805 (1979)
45. Martinec, E.J., Sotkovab, G.M.: Superghosts revisited: supersymmetric bosonization. *Phys. Lett. B* **208**, 249–254 (1988)
46. Verlinde, E.P., Verlinde, H.L.: Lectures on string perturbation theory, preprint IASSNS-HEP-88-52
47. Bump, D.: Lie Groups. Graduate Text in Mathematics, vol. 225. Springer, Berlin (2004)
48. Greene, R.E., Krantz, S.G.: Function Theory of One Complex Variable. Graduate Studies in Mathematics, vol. 40. AMS (2006)
49. Allison, B., Azam, S., Berman, S., Gao, Y., Pianizola, A.: Extended affine Lie Algebras and their root system. *Mem. Am. Math. Soc.* **126** (1997)
50. Neher, E.: Lectures on Extended Affine Lie Algebras. Fields Institute Communications, vol. 59, pp. 107–167. American Mathematical Society, Providence, RI (2011). [arXiv:1003.2352](https://arxiv.org/abs/1003.2352)
51. Slodowy, P.: Beyond Kac-Moody algebras and inside. In: Canadian Mathematical Society, Conference Proceedings, vol. 5, pp. 361–371 (1986)
52. Kac, V.G.: Infinite Dimensional Lie Algebras. Cambridge University Press (1990)
53. Goddard, P., Olive, D. (eds.), Kac-Moody and Virasoro Algebras: A Reprint Volume for Physicists. Advanced Series in Mathematical Physics, vol. 3. World Scientific (1988)
54. Serre, J.-P.: Complex Semisimple Lie Algebras. Springer, Berlin (2001)
55. Fuchs, J., Schweigert, C.: Symmetries, Lie Algebras and Representations. A Graduate Course for Physicists. Cambridge Monographs on Mathematical Physics, Cambridge University Press (1997)
56. Bourbaki, N.: Éléments de mathématique. Groupes et algèbres de Lie. Chapitre 4 à 6. Springer, Berlin (2007)
57. Stekolshchik, R.: Notes on Coxeter Transformations and the McKay correspondence. Springer Monographs in Mathematics. Springer, Berlin (2008)
58. Knizhnik, V.G., Zamolodchikov, A.B.: Current algebra and Wess-Zumino models in two dimensions. *Nucl. Phys. B* **247**, 83–103 (1984)
59. Kohno, T.: Hecke algebra representations of braid groups and classical Yang-Baxter equations. In: Conformal Field Theory and Solvable Lattice Models. Advanced Studies in Pure Mathematics, vol. 16, pp. 225–269 (1988)
60. Kohno, T.: Monodromy representations of braid groups and Yang-Baxter equations. *Ann. Inst. Fourier* **37**, 139–160 (1987)
61. Kassel, C.: Quantum Groups. Graduate Texts in Mathematics, vol. 155. Springer, Berlin (1995)
62. Baxter, R.J.: Exactly Solved Models in Statistical Mechanics. Academic Press (1982)
63. Arnol'd, V.I.: The cohomology ring of the colored braid group. *Math. Notes* **5**, 138–140 (1969)
64. Kac, V.G., Frenkel, I.B.: Basic representations of affine Lie algebras and dual resonance models. *Inven. Math.* **62**, 23–66 (1980/81)
65. Segal, G.: Unitary representations of some infinite dimensional groups. *Commun. Math. Phys.* **80**, 301 (1981)
66. Goddard, P., Nahm, W., Olive, D., Schwimmer, A.: Vertex operators for non-simply-laced algebras. *Commun. Math. Phys.* **107**, 179–212 (1986)
67. McCoy, B.M., Wu, T.T.: The Two-Dimensional Ising Model. Harvard University Press (1973)
68. Itzykson, C., Drouffe, J.-M.: Statistical field theory. In: From Brownian Motion to Renormalization and Lattice Gauge Theory, vol. 1. Cambridge University Press (1989)
69. Witten, E.: Constraints on supersymmetry breaking. *Nucl. Phys. B* **202**, 253–316 (1982)
70. Gliozzi, F., Scherk, J., Olive, D.: Supersymmetry, supergravity theories, and the dual spinor model. *Nucl. Phys. B* **122**, 253 (1977)

71. Postnikov, M.M.: Lectures in Geometry: Lie Groups and Lie Algebras. Editorial URSS (1994)
72. Serre, J.-P.: A Course in Arithmetics. Springer, Berlin (1973)
73. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices and Groups, 3rd edn. Springer, Berlin (1999)
74. Jordan, P., Wigner, E.: Über das paulische äquivalenzverbot. Zeitschrift für Physik **47**, 631 (1928)
75. Kostelecky, V.A., Lechtenfeld, O., Lerche, W., Samuel, S., Watamura, S.: Conformal techniques, bosonization and tree level string amplitudes. Nucl. Phys. B **288**, 173–232 (1987)
76. Green, M.B., Schwarz, J.H.: Covariant deflist of superstrings. Phys. Lett. B **136**, 367 (1984)
77. Adams, J.F.: Lectures on Exceptional Lie Groups. University of Chicago Press (1996)
78. Mumford, D.: Tata Lectures on Theta. Modern Birkhäuser Classics, vols. I, II, III. Springer, Berlin (2007)
79. Bouwknegt, P., Schoutens, K.: W symmetry in conformal field theory. Phys. Rep. **223**, 183–276 (1993). [arXiv:hep-th/9210010](https://arxiv.org/abs/hep-th/9210010)
80. Pope, C.N.: W strings 93. [arXiv:hep-th/9309125](https://arxiv.org/abs/hep-th/9309125)
81. Hull, C.M.: Lectures on W gravity, W geometry and W strings. [arXiv:hep-th/9302110](https://arxiv.org/abs/hep-th/9302110)
82. Tye, S.H.H.: Status of fractional superstrings. [arXiv:hep-th/9311021](https://arxiv.org/abs/hep-th/9311021)
83. Bröcker, T., Tom Dieck, T.: Representations of Compact Lie Groups. Springer, Berlin (1985)
84. Kac, V.G.: Classification of simple Lie superalgebras. Funct. Anal. Appl. **9**, 263–265 (1975)
85. Witten, E.: Nonabelian bosonization in two dimensions. Commun. Math. Phys. **92**, 455 (1984)
86. Polyakov, A.M., Wiegmann, P.B.: Goldstone fields in two-dimensions with multivalued actions. Phys. Lett. B **141**, 223 (1984)
87. Wess, J., Zumino, B.: Consequences of anomalous ward identities. Phys. Lett. **37**, 95 (1971)
88. Di Vecchia, P., Durhuus, B., Petersen, J.L.: The Wess-Zumino action in two-dimensions and nonabelian bosonization. Phys. Lett. B **144**, 245 (1984)
89. Gantmacher, F.: The Theory of Matrices. AMS (2000)
90. Hazewinkel, M., Hesseling, W., Siersma, J.D., Veldkamp, F.: The ubiquity of Coxeter Dynkin diagrams. (An introduction of the A-D-E problem), Nieuw Archief v. Wiskunde **35**, 257–307 (1977)
91. Durfee, A.: Fifteen characterizations of rational double points and simple critical points. Enseign. Math. **25**, 131–163 (1979)
92. Capelli, A., Zuber, J.B.: A-D-E classification of conformal field theories. Scholarpedia **5**(4), 10314 (2010). [arXiv:0911.3242](https://arxiv.org/abs/0911.3242)
93. Kodaira, K.: On compact analytic surfaces, II. Ann. Math. **77**, 563–626 (1963)
94. Kodaira, K.: On compact analytic surfaces, III. Ann. Math. **78**, 1–40 (1963)
95. Kodaira, K.: On the structure of compact complex analytic surfaces, I. Am. J. Math. **86**, 751–798 (1964)
96. Capelli, A., Itzykson, C., Zuber, J.-B.: Modular invariant partition functions in two dimensions. Nucl. Phys. **B280**[FS 18], 445–465 (1987)
97. Lerche, W., Vafa, C., Warner, N.P.: Chiral rings in $N = 2$ superconformal theories. Nucl. Phys. B **324**, 427 (1989)
98. Martinec, E.J.: Algebraic geometry and effective Lagrangians. Phys. Lett. B **217**, 431 (1989)
99. McKay, J.: Graphs, singularities, and finite groups. In: The Santa Cruz Conference on Finite Groups, Proceedings of Symposia in Pure Mathematics, vol. 37, pp. 183–186. AMS (1979, 1980)
100. Peterfalvi, T.: Character Theory for the Odd Order Theorem. London Mathematical Society Lecture Notes Series, vol. 272. Cambridge University Press (2000)
101. Lang, S.: Algebra. Springer, Berlin (2002)
102. Serre, J.-P.: Linear Representations of Finite Groups. Springer, Berlin (1977)
103. Arnold, V.I., Gusein-Zade, S.M., Varchenko, A.N.: Singularities of Differential Maps, vols. I, II. Birkhäuser (1985, 1988)
104. Reid, M.: The Du Val singularities A_n , D_n , E_6 , E_7 , E_8 . <https://homepages.warwick.ac.uk/~masda/surf/more/DuVal.pdf>

105. Reid, M.: La correspondance de McKay, Séminaire Bourbaki 1999/2000. *Astérisque* **276**, 53–72 (2002)
106. Barth, W., Hulek, K., Peter, C., van de Ven, A.: *Compact Complex Surfaces*. *Ergebnisse der Mathematik und ihrer Grenzgebiete*, 3. Folge, 2nd edn. Band 4. Springer, Berlin (2004)
107. Klein, F.: *Vorlesungen über das Icosaeder und die Auflösung der Gleichung vom fünften Grade*. Teubner, Leipzig (1884)
108. Milnor, J.: *Singular Points of Complex Hypersurfaces*. *Annals of Math Studies*, vol. 61. Princeton University Press (1968)

Part II

Constructing Superstring Theory

Part II consists of five chapters.

Chapter 3 describes the quantization of the superstring in the old covariant formalism as well as from the modern BRST perspective. BRST-invariant physical vertices are discussed and constructed using bosonization techniques. Some general properties of the physical amplitudes are deduced, in particular, the spacetime SUSY Ward identities. The proof of the non-ghost theorem is sketched with more details in the **Appendix**.

Chapter 4 describes the techniques to compute quantum amplitudes in the bosonic string theory as a preparation for the superstring.

Chapter 5 gives the construction of all consistent string theories moving in flat 10d whose world-sheet theory has local $(1,1)$ superconformal symmetries. The condition for modular invariance and absence of 2d global gravitational anomalies in all genera is described in detail. The partition functions of the various superstring theories is computed. For the open string, the condition for the cancellation of tadpoles is deduced by a direct computation of the relevant quantum amplitudes.

In Chap. 6, we discuss T-duality, D-branes, orbifolds, and orientifolds in the context of the bosonic string.

In Chap. 7, we construct the heterotic string in both the fermionic and bosonic formulations. Toroidal compactifications and their physics are described in detail. The BPS states are also described.

Chapter 3

Spectrum, Vertices, and BRST Quantization



Abstract We present the modern covariant quantization of perturbative string theory *à la* BRST. We determine the physical spectrum of the several string theories, write down the vertices of physical states, and prove unitarity. We discuss the role of the picture charge in the superstring, and deduce several general properties of perturbative (super)string theory. In the three preliminary sections we discuss subtle points of bosonization, light-cone quantization, and the old covariant quantization. In the last section, the Chan–Paton degrees of freedom are introduced and studied.

With conformal tools at our disposal, we may proceed to quantize the bosonic string and the superstring. There are several distinct string theories we may consider. For the moment, we limit ourselves to the simplest models leaving more general constructions for later chapters. Except in Sect. 3.9 all strings are closed and oriented.

Our first task is to determine the physical spectrum of a string theory. We may use different formalisms. The first one in historical order is *light-cone quantization*: as mentioned in Chap. 1, it is manifestly unitary but not Lorentz covariant. The second one is Old Covariant Quantization (OCQ) which is covariant but ad hoc. The modern intrinsic method is BRST quantization which is fully covariant and canonical.

By the CFT state-to-operator map, for each physical state we have an operator called its *vertex*. BRST cohomology on operators then determines the physical states. In a covariant quantization, the vertices carry a representation of the space-time Lorentz group. From the world-sheet perspective, this symmetry becomes a 2d current algebra. We start with some preliminary consideration on this crucial topic.

3.1 The Superstring Lorentz Current Algebra

In the context of the superstring with gauge-fixed world-sheet action

$$\frac{1}{4\pi} \int \left(\partial X^\mu \bar{\partial} X_\mu + \psi^\mu \bar{\partial} \psi_\mu + \tilde{\psi}^\mu \partial \tilde{\psi}_\mu + \text{ghosts} \right), \quad (3.1)$$

we continue the discussion of the Lorentz current algebra of the fermions ψ^μ , following (2.490)–(2.493). There we saw that the $\text{Spin}(d)$ spin fields $S_\alpha(z)$ are local

with respect to themselves iff 8 divides d . This may seem odd, since the superstring critical dimension, 10, is not a multiple of 8. Indeed that discussion—valid for a *unitary* current algebra—is directly applicable in the light-cone gauge, where the world-sheet theory is a 2d unitary SCFT involving only the transverse fermions ψ^i , $i = 1, 2, \dots, d_{\text{crit}} - 2 \equiv 8$. In a covariant gauge, the 2d CFT is *non-unitary* and additional subtleties play a role.

We are interested in the full covariant $SO(9, 1)$ current algebra of the superstring or, rather, its Wick rotated $SO(10)$ version. A spin field $S_\alpha(z)$ maps the NS sector into the R one, so the supercurrent $T_F(w)$ is doubly valued around $S_\alpha(z)$. Consistency¹ requires the ghost fields $\beta(w)$, $\gamma(w)$ to be also doubly valued around the spin field. This implies that the matter system spin field $S_\alpha(z)$ must be accompanied by a spin field of the β , γ system. Physical spin fields contain both matter- and ghost-sector spin operators. It is convenient to adopt the language of bosonization from Chap. 2.

In the bosonized setup, we unify the scalar $\phi_g \equiv \phi$ which bosonizes the anomalous ghost $U(1)$ current $-\beta\gamma$ with the five scalars ϕ_i ($i = 1, \dots, 5$) which bosonize the Cartan currents of $SO(10)_1$. In this unified system, we have the *six* Cartan currents

$$i\partial\phi_1, i\partial\phi_2, i\partial\phi_3, i\partial\phi_4, i\partial\phi_5, \partial\phi_g, \quad (3.2)$$

$$\text{with OPE } \phi_a(z)\phi(w) \sim -\delta_{a,b} \log(z-w), \quad a, b = 1, 2, 3, 4, 5, g. \quad (3.3)$$

In Eq. (3.2) the $U(1)$ ghost current, $\partial\phi_g$, has no i since it has opposite Hermitian properties² with respect to the matter currents. For our purposes it is convenient to derogate from standard conventions, and write $\phi_g \equiv \phi = i\phi_6$. Now the OPE reads

$$\begin{aligned} \phi_a(z)\phi_b(w) &\sim -\eta_{ab} \log(z-w), \\ \text{where } \eta_{ab} &= \text{diag}(+1, +1, +1, +1, +1, -1). \end{aligned} \quad (3.4)$$

The six currents $i\partial\phi_a$ plus the 60 currents associated with the roots³

$$c_{\pm e_a \pm e_b} e^{\pm i\phi_a \pm i\phi_b} \quad a \neq b \quad (3.5)$$

generate a $SO(10, 2)$ current algebra (which is just a *subalgebra* of the actual current algebra; see below). More generally, we consider primary operators of the form

$$O_\lambda(z) = c_\lambda e^{i\lambda \cdot \phi(z)} \quad (3.6)$$

where now the weight

$$\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5; \lambda_6) \quad (3.7)$$

¹ γ is a wrong-statistics SUSY parameter and so has the same spin-structure as T_F . Alternatively: β , γ should carry the same spin-structure as T_F for the BRST current to be a 1-form on the world-sheet.

² For ordinary 2d fermions, one has $(\lambda^{e_j})^\dagger = \lambda^{-e_j}$, so that Hermitian conjugation flips the sign of the $U(1)$ charge of “matter” fermions, while for ghost bosons $\gamma^\dagger = \gamma$, and hence Hermitian conjugation leaves the $U(1)$ charge of bosonic ghosts unchanged.

³ $c_{\pm e_a \pm e_b}$ is a suitable cocycle, whose form is similar to the one for the $SO(8n)$ case [1].

takes value in a *Lorentzian* lattice. We write $\mathbb{Z}^{5,1}$ for the standard Lorentzian lattice of signature (5, 1), i.e. $\mathbb{Z}^6$ endowed with the *indefinite* inner product

$$\lambda \cdot \lambda' = \eta^{ab} \lambda_a \lambda'_b \equiv \sum_{j=1}^5 \lambda_j \lambda'_j - \lambda_6 \lambda'_6. \quad (3.8)$$

The weights of the $\mathfrak{so}(10, 2)$ root operators (3.5) generate the root sublattice

$$\Lambda_{\text{root}} = \{ \lambda \in \mathbb{Z}^{5,1} \mid \iota \cdot \lambda \in 2\mathbb{Z} \} \subset \mathbb{Z}^{5,1}, \quad \text{where } \iota = (1, 1, 1, 1, 1; 1). \quad (3.9)$$

Again, the weight lattice is the dual one

$$\Lambda_{\text{weight}} = \Lambda_{\text{root}}^\vee = \{ \lambda \in \tfrac{1}{2}\mathbb{Z}^{5,1} \mid \lambda_i - \lambda_j = 0 \pmod{1} \}. \quad (3.10)$$

All these lattices are endowed with the Lorentzian quadratic form (3.8).

An operator $\mathcal{O}_\lambda(z)$ is local relatively to the $SO(10, 2)$ currents (3.5) if and only if $\lambda \in \Lambda_{\text{weight}}$, so, again, Λ_{weight} parametrizes the operators of the form (3.6) which may be present in a local CFT containing the $SO(10, 2)$ currents.

The *Fermi parity* (a.k.a. *chirality*) of the operator $\mathcal{O}_\lambda(z)$ is $(-1)^{\iota \cdot \lambda}$, i.e.

$$(-1)^F \mathcal{O}_\lambda(z) (-1)^F = (-1)^{\iota \cdot \lambda} \mathcal{O}_\lambda(z). \quad (3.11)$$

One has the OPEs⁴ (cf. Eq. (2.510))

$$\mathcal{O}_\lambda(z) \mathcal{O}_{\lambda'}(w) = (z - w)^{\lambda \cdot \lambda'} \varepsilon(\lambda, \lambda') e^{i\lambda \cdot \phi(z) + i\lambda' \cdot \phi(w)} c_{\lambda + \lambda'}. \quad (3.12)$$

Again the quotient $\Lambda_{\text{weight}}/\Lambda_{\text{root}}$ consists of the four classes

$$(o), \quad (v), \quad (s), \quad (c), \quad (3.13)$$

and all considerations in Sect. 2.9 apply word-for-word to the present context

$$(v) \cdot (v) \in \mathbb{Z}, \quad (s) \cdot (s) \in \mathbb{Z}, \quad (c) \cdot (c) \in \mathbb{Z} \quad (3.14)$$

$$(v) \cdot (s) \in \mathbb{Z} + \tfrac{1}{2}, \quad (v) \cdot (c) \in \mathbb{Z} + \tfrac{1}{2}, \quad (s) \cdot (c) \in \mathbb{Z} + \tfrac{1}{2}. \quad (3.15)$$

In particular, the full covariant spin operators

⁴ Here $\varepsilon(\lambda, \lambda')$ is the *cocycle* in the math sense of the term, which satisfies the cocycle condition

$$\varepsilon(\lambda, \lambda') \varepsilon(\lambda + \lambda', \lambda'') + \varepsilon(\lambda', \lambda'') \varepsilon(\lambda' + \lambda'', \lambda) + \varepsilon(\lambda'', \lambda) \varepsilon(\lambda'' + \lambda, \lambda') = 0.$$

For λ, λ' positive roots $\varepsilon(\lambda, \lambda')$ reduce to the Lie algebra structure constants in the Serre basis $[E_\lambda, E_{\lambda'}] = \varepsilon(\lambda, \lambda') E_{\lambda + \lambda'}$ (thus the cocycle condition is just the Jacobi identity). $\varepsilon(\lambda, \lambda')$ is related to the c_λ by $c_\lambda \cdot c_{\lambda'} = \varepsilon(\lambda, \lambda') c_{\lambda + \lambda'}$.

$$O_\lambda(z) \quad \text{with } \lambda \in (s) \text{ or } (c) \quad (3.16)$$

are local with respect to themselves, as they should. The GSO projections work exactly as in the $SO(8n)$ case, that is, to get a *local* operator algebra containing the current algebra and a spin field we must keep either the two operator classes (o) and (s) , or (o) and (c) . These two projections correspond, respectively, to keeping only the operators $O_\lambda(z)$ with weights in one of the two *self-dual* integral sublattices

$$\Lambda_{(s)} = \{\lambda \in \Lambda_{\text{weight}} \mid \sigma_+ \cdot \lambda \in 2\mathbb{Z}\}, \quad \Lambda_{(c)} = \{\lambda \in \Lambda_{\text{weight}} \mid \sigma_- \cdot \lambda \in 2\mathbb{Z}\}, \quad (3.17)$$

where $\sigma_\pm = (1, 1, 1, 1, 1; \pm 1)$. In both cases, the operator class (v) is projected out. As we shall see, this eliminates the tachyon from the physical spectrum.

Our system may look similar to an $SO(12)$ current algebra, but this is *not* correct: the present situation is radically different from the one discussed in Sect. 2.9. In the $SO(8n)$ case, the lattice Γ_{root} was positive-definite, and there were only finitely many lattice elements of square-length 2; in fact, they were precisely the $4n(8n - 2)$ roots of the finite-dimensional Lie algebra $\mathfrak{so}(8n)$. On the contrary in the Lorentzian lattice Λ_{root} (resp. Λ_{weight}), there are *infinitely* many elements of square-length 2 (resp. of any given square-length). This infinity reflects two important physical facts:

- (i) the ghost system β, γ is *not* a unitary CFT. In particular, its spectrum of conformal weights h is unbounded below;
- (ii) as discussed in Sect. 2.5.7, each physical state should appear in the Hilbert space *infinitely many times* at different picture levels. By the CFT state-operator correspondence, the same should hold for the vertices which have copies in all picture levels in $\mathbb{Z} + q$, with $q = 0$ in the NS sector and $q = \frac{1}{2}$ in the R-sector.

The operator $O_\lambda(z)$ has picture charge $q \equiv \lambda_6$. Clearly in Λ_{weight} , there are only finitely many elements with a given square-length $\lambda \cdot \lambda$ and a given λ_6 . The conformal weight of the operator $O_\lambda(z)$ is

$$h_\lambda = \frac{1}{2}\lambda \cdot \lambda - \lambda_6, \quad (3.18)$$

where the linear term is the contribution from the background charge of the ghost current algebra: to get (3.18) we used Eq. (2.270) and the fact that the superconformal ghost β, γ system has $\lambda = \frac{3}{2}$, hence $Q = +2$. The spin fields in class (s) (or (c)) have half-integral picture charge λ_6 . The basic spin fields have weights

$$\left(\pm \frac{1}{2}, \pm \frac{1}{2}, \pm \frac{1}{2}, \pm \frac{1}{2}, \pm \frac{1}{2}; \pm \frac{1}{2} \right) \quad (3.19)$$

with picture charge $\lambda_6 = \pm \frac{1}{2}$. Their dimension is

$$h_\pm = \frac{1}{2} \mp \frac{1}{2}. \quad (3.20)$$

The spin field of picture $-\frac{1}{2}$ has the right dimension, $h_- = 1$, for its integral on Σ to be conformal invariant: this is a necessary condition for a physical vertex. It is therefore a natural candidate for the vertex of a physical massless spacetime fermion. We shall see that this expectation is correct in Sect. 3.5. The GSO projection keeps only spin operators in (s) or in (c) . In the first case, the $SO(10)$ part of the $\lambda_6 = -\frac{1}{2}$ surviving spin operator has a weight (3.19) with an odd number of $-\frac{1}{2}$'s

$$\lambda = \left(\overbrace{\pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}, \pm\frac{1}{2}}^{\text{odd \# of } -\frac{1}{2}}; -\frac{1}{2} \right), \quad (3.21)$$

i.e. it has *negative* $SO(10)$ chirality. Thus our candidate (left-moving⁵) spinorial vertex has the form⁶

$$e^{-\phi(z)/2} S_{\dot{\alpha}}(z) e^{ik \cdot X(z)}, \quad k^2 = 0, \quad (3.22)$$

with $S_{\dot{\alpha}}$ a negative chirality $SO(10)$ spin field. The $SO(10)$ part of the other spin field in class (s) has picture charge $+\frac{1}{2}$, $h_+ = 0$, and positive $SO(10)$ chirality, namely it has the form

$$e^{\phi(z)/2} S_{\alpha}(z). \quad (3.23)$$

The physical isomorphism between different picture levels, whose existence was suggested in Sect. 2.5.7, requires the existence of a picture $+\frac{1}{2}$ version of the massless fermion vertex in addition to the picture $-\frac{1}{2}$ one (3.22). The $+\frac{1}{2}$ massless fermion vertex then must have the general form

$$e^{\phi(z)/2} S^{\alpha}(z) \left(\text{a } h = 1 \text{ operator in the } X^{\mu}, \psi^{\mu} \text{ system} \right)_{\alpha\dot{\alpha}} e^{ik \cdot X(z)}. \quad (3.24)$$

The unknown operator in the parenthesis then must have the form

$$(a \partial X_{\mu} + b \psi_{\mu} k \cdot \psi) (\gamma^{\mu})_{\alpha\dot{\alpha}}, \quad (3.25)$$

for some coefficients a, b . The BRST analysis will confirm these conclusions and determine a, b . In the same way, the $h = 1$ class (o) operators with picture -1 are $SO(10)$ vectors, and hence we expect massless vector vertices of the form

$$e^{-\phi(z)} \psi^{\mu}(z) e^{ik \cdot X(z)}, \quad k^2 = 0. \quad (3.26)$$

⁵ The full physical vertex contains in addition a right-moving factor with $\tilde{h} = 1$.

⁶ The on-shell condition $k^2 = 0$ that says that the corresponding state (if physical) is massless follows from the fact that the full vertex (3.22) should have $h = 1$, while $h(e^{ik \cdot X}) = \frac{1}{2}k^2$.

3.2 The Physical Spectrum: Light-Cone Gauge

The simplest approach to the spectrum is the light-cone gauge, in which unitarity (i.e. positive-definiteness of the Hilbert space norm) is manifest, while Poincaré invariance is not. The existence of a gauge where unitarity is manifest is conceptually important since it entails that gauge-fixing independence implies unitarity, i.e. we may prove unitarity simply by performing the gauge transformation to the light-cone gauge. This argument requires the gauge symmetry to be non-anomalous, i.e. the string to be in critical dimension.

We write the formulae for the superstring; the expressions for the bosonic string are obtained from them by omitting the world-sheet spinorial fields (see also Sect. 1.3).

In the light-cone gauge the 2d dynamical d.o.f. are the matter fields X^i , ψ^i , and $\tilde{\psi}^i$ in the transverse directions⁷

$$S_{\text{trans.}} = \frac{1}{4\pi} \int d^2z \left(\partial X^i \bar{\partial} X^i + \psi^i \bar{\partial} \psi^i + \tilde{\psi}^i \partial \tilde{\psi}^i \right), \quad (3.27)$$

where $i = 1, \dots, d-2$. The fields X^+ and ψ^+ are fixed by the gauge conditions⁸

$$X^+ = \alpha_0^+ \tau, \quad \psi^+ = 0, \quad (3.28)$$

and the fields ∂X^- , ψ^- are determined in terms of the transverse d.o.f. by solving the constraints $L_n = G_r = 0$.

! Physical states beyond the reach of light-cone quantization

The gauge (3.28) makes sense only when $\alpha_0^+ \neq 0$, i.e. $p^+ \neq 0$. For states with $p_\mu \neq 0$ we can find a frame with $p^+ \neq 0$. Thus the light-cone gauge allows to study all physical states with $p_\mu \neq 0$, but states which *exist only* at $p_\mu = 0$ are out of its reach.

The mass-shell condition for physical states of the open superstring is

$$(L_0 - a)|\text{phys}\rangle = 0, \quad (3.29)$$

where the shift a is different for the bosonic and the superstring and in the second case it depends on the sector NS versus R. The shift a vanishes in the R-sector since

⁷ Equation (3.27) holds with the convention $\alpha' = 2$ otherwise the pre-factor reads $1/2\pi\alpha'$.

⁸ We recall that $\alpha_0^+ = (2\alpha')^{1/2} p^+$ for the open string, and $\alpha_0^+ = (\alpha'/2)^{1/2} p^+$ for the closed one.

the fermionic and bosonic ground energies cancel each other by 2d SUSY.⁹ In the NS sector, a is easily computed by ζ -function regulation to be (cf. **BOX 1.2**)

$$a = (d - 2) \left(\frac{1}{24} + \frac{1}{48} \right) \equiv \frac{(d - 2)}{16}, \quad (3.30)$$

while in the bosonic case we omit the fermion contribution to the zero-point energy

$$(\text{bosonic string}) \quad a = \frac{(d - 2)}{24}. \quad (3.31)$$

The mass-square of an open superstring state moving in flat $\mathbb{R}^{d-1,1}$ space is then

$$2m^2 = N_{\text{l.c.}} - \frac{d - 2}{16} \quad (\text{open superstring}) \quad (3.32)$$

where $N_{\text{l.c.}}$ is the *light-cone level*

$$N_{\text{l.c.}} \equiv \sum_{n>0} \alpha_{-n}^i \alpha_n^i + \sum_{r \geq 0} r \psi_{-r}^i \psi_r^i. \quad (3.33)$$

It is easy to see that spacetime Poincaré invariance implies $d = 10$. Indeed consider the states in the first excited level of the NS sector: there are just $d - 2$ of them

$$\psi_{-1/2}^i |0; p_\mu\rangle_{\text{NS}}, \quad i = 1, 2, \dots, d - 2, \quad (3.34)$$

forming the vector representation of $SO(d - 2)$. Massive states form representations of the larger spin group $SO(d - 1)$. Thus the states (3.34) cannot correspond to a massive state, and hence they must be massless. Then, from (3.32),

$$0 = \frac{1}{2} - \frac{d - 2}{16} \Rightarrow d = 10. \quad (3.35)$$

The analogue argument for the bosonic string gives $d = 26$; see Sect. 1.3.

In the *closed* superstring, we have two copies of the above story

$$(L_0 - a)|\text{phys}\rangle = 0, \quad (\tilde{L}_0 - a)|\text{phys}\rangle = 0, \quad (3.36)$$

so (keeping track of factors 2 in the respective mode expansions of ∂X^μ) the mass condition becomes

$$\frac{1}{2} m^2 = N_{\text{l.c.}} - \nu = \tilde{N}_{\text{l.c.}} - \tilde{\nu} \quad (\text{closed superstring}) \quad (3.37)$$

⁹ This is the historic naive argument. The deep reasoning goes as follows: a transverse CFT Ramond vacuum has $h = \hat{c}/16 = 1/2$. Then “ L_0 ” defined by “normal ordering” the oscillators so that L_0 vanishes on the Ramond vacuum which is actually $L_0 - 1/2$ in terms of L_0 as defined in CFT.

($\nu, \tilde{\nu} = 0, \frac{1}{2}$ for R resp. NS). Equation (3.37) implies the right–left matching condition

$$N_{\text{l.c.}} - \nu = \tilde{N}_{\text{l.c.}} - \tilde{\nu}. \quad (3.38)$$

To get the superstring, one has to enforce the GSO projection which keeps only the states such that

$$(-1)^F |\text{phys}\rangle = -|\text{phys}\rangle. \quad (3.39)$$

The fermion parity $(-1)^F$ acts as the $\text{Spin}(8)$ chirality matrix Γ_9 on the Ramond ground states which in light-cone quantization form a $SO(8)$ Majorana spinor. It is not obvious that this projection is consistent with 2d locality and string interactions; however this turns out *to be true* as it will be clear from the covariant quantization.

Examples: first levels of the open string

NS sector

- at level $-\frac{1}{2}$, we have the vacuum $|0\rangle$ of the transverse SCFT (3.27). The on-shell state,

$$|0; p\rangle \equiv e^{ip_\mu X^\mu(0)} |0\rangle, \quad p^2 = \frac{1}{4}, \quad (3.40)$$

is a *tachyon* which is happily projected out by the GSO projection. This eliminates the problem with the stability of the perturbative vacuum we had in the bosonic string;

- at level 0 we have the state $\psi_{-1/2}^i |0; p\rangle$, a massless vector in the $\mathbf{8}_v$ of the little group $SO(8)$;

BOX 3.1 - The representation ring of $\text{Spin}(2n)$

We write Λ_k ($k = 0, 1, \dots, 2n$) for the k -index totally antisymmetric representation of $SO(2n)$, and S_+, S_- for the (irreducible) spin representations of chirality $+$ and $-$, respectively. For $k \neq n$ the representation Λ_k is irreducible and $\Lambda_k \simeq \Lambda_{2n-k}$, while Λ_n splits into self-dual and anti-self-dual parts, $\Lambda_n = \Lambda_n^+ \oplus \Lambda_n^-$. $\mathbf{Sy}$ stands for the 2-index symmetric traceless representation

Theorem 3.1 (*e.g. Sect. VI.(6.2) [2]*) *The representation ring of $\text{Spin}(2n)$ is the polynomial ring*

$$\text{RSpin}(2n) = \mathbb{Z}[\Lambda_1, \dots, \Lambda_{n-2}, S_+, S_-]$$

and one has

$$\begin{aligned} S_+ \otimes S_+ &= \Lambda_n^+ \oplus \Lambda_{n-2} \oplus \Lambda_{n-4} \oplus \dots & S_- \otimes S_- &= \Lambda_n^- \oplus \Lambda_{n-2} \oplus \Lambda_{n-4} \oplus \dots \\ S_+ \otimes S_- &= \Lambda_{n-1} \oplus \Lambda_{n-3} \oplus \Lambda_{n-5} \oplus \dots & \Lambda_1 \otimes \Lambda_1 &= \Lambda_0 \oplus \Lambda_2 \oplus \mathbf{Sy}. \end{aligned}$$

In particular, for $\text{Spin}(4n)$ (resp. $\text{Spin}(4n+2)$) the tensor product of two irreducible spin representations of the same chirality is the direct sum of the spaces of *even* (resp. *odd*) forms subjected to a self-dual or anti-self-dual condition as in the statement of the **theorem**.

- at level $+\frac{1}{2}$ the states $\alpha_{-1}^i |p\rangle$ and $\psi_{-1/2}^i \psi_{-1/2}^j |p\rangle$ with positive mass in the $\mathbf{8}_v \oplus \mathbf{28}$ representations of $SO(8)$. These states are *projected out* by GSO.
- at level $+1$ we have $\psi_{-1/2}^i \psi_{-1/2}^j \psi_{-1/2}^k |p\rangle$ in the $\mathbf{56}$ of $SO(8)$, $\alpha_{-1}^i \psi_{-1/2}^j |p\rangle$ in the $\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}$, and $\psi_{-3/2}^i |p\rangle$ in the $\mathbf{8}_v$. Lorentz symmetry requires these state representations to form complete $SO(9)$ representations. The natural ones are the 3-index antisymmetric $\mathbf{84}$ and the two-index symmetric traceless $\mathbf{44}$. Under the subgroup $SO(8) \subset SO(9)$ one has $\mathbf{84} = \mathbf{56} \oplus \mathbf{28}$, and $\mathbf{44} = \mathbf{35} \oplus \mathbf{8}_v \oplus \mathbf{1}$, and we have full agreement. These states are kept by GSO.

R-sector

- at level 0 the ground states $|\alpha; p_\mu\rangle$ in the spinor $\mathbf{8}_s$ irrepr. of $SO(8)$ of chirality -1 (in the present conventions) and $|\dot{\alpha}; p_\mu\rangle$ in the spinor $\mathbf{8}_c$ irrepr. of $SO(8)$ of chirality $+1$. Only the first (or the second) is kept by GSO. Since it is a representation of $SO(8)$ which cannot be extended to a representation of $SO(9)$, it should be massless;
- at level $+1$ the GSO allowed states $\alpha_{-1}^i |\alpha; p_\mu\rangle$ in the $\mathbf{8}_c \oplus \mathbf{56}_c$ of $SO(8)$ and $\psi_{-1}^i |\dot{\alpha}; p_\mu\rangle$ in the $\mathbf{8}_s \oplus \mathbf{56}_s$; together they form the $\mathbf{128}$ of $SO(9)$ (i.e. the γ -traceless part of the vector-spinor: $\Psi_{\mu\alpha}$ with $(\gamma^\mu)^{\alpha\beta} \Psi_{\mu\beta} = 0$). In addition we have GSO forbidden states of opposite chirality.

The *closed* superstring Hilbert space is the tensor product of two copies of the open Hilbert space, subject to the level-matching constraint (3.38). The GSO projections on the left and on the right are independent. Thus (up to conventional choices) we have two possibilities: either we use the *same* GSO projection on the two sides or we use opposite ones (cf. Sect. 2.9). The two possibilities correspond to two different superstring theories which, when classifying possible 10d superstring theories in Chap. 5, we shall call, respectively, Type IIB and Type IIA.

We write $R_\pm$ (resp. $NS_\pm$) for the Ramond (Neveu–Schwarz) sector with GSO projection ∓ 1 . Note that only NS_- has half-integral level, so the constraint (3.38) allows only its pairing with itself, (NS_-, NS_-) . We have seen¹⁰ that this sector should be projected out by locality as long as we have a $(R_\pm, *)$ sector; we will show in Sect. 5.1 that its presence is incompatible with the absence of 2d gravitational anomalies.¹¹ This is fortunate, since the projected out sector (NS_-, NS_-) contains a tachyon; thus the physical spectrum of the consistent superstring is tachyon-free. All other sectors consist of massless particles plus an infinite tower of massive states.

The $SO(8)$ representation content of the massless sectors is determined using the group-theoretic facts in BOX 3.1: see Table 3.1. From the table, we see that the

Table 3.1 $SO(8)$ representation content of the massless sectors in the closed superstring. Λ_k stands for the k -index antisymmetric representation, Sy stands for the 2-index symmetric

Sector	$SO(8)$ representation	Dimension
(NS_+, NS_+)	$\mathbf{8}_v \otimes \mathbf{8}_v = \Lambda_0 \oplus \Lambda_2 \oplus Sy$	$\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}$
(R_+, R_+)	$\mathbf{8}_s \otimes \mathbf{8}_s = \Lambda_0 \oplus \Lambda_2 \oplus \Lambda_4^+$	$\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}_+$
(R_-, R_-)	$\mathbf{8}_c \otimes \mathbf{8}_c = \Lambda_0 \oplus \Lambda_2 \oplus \Lambda_4^-$	$\mathbf{1} \oplus \mathbf{28} \oplus \mathbf{35}_-$
(NS_+, R_+)	$\mathbf{8}_v \otimes \mathbf{8}_s$	$\mathbf{8}_c \oplus \mathbf{56}_s$
(NS_+, R_-)	$\mathbf{8}_v \otimes \mathbf{8}_c$	$\mathbf{8}_s \oplus \mathbf{56}_c$

¹⁰ Cf. Sect. 3.1. The analysis of locality will be more clear below in the OCQ approach.

¹¹ Assuming the presence of spacetime fermions in the physical spectrum.

massless states in the NS-NS sector are as in the bosonic string: the metric $G_{\mu\nu}(\mathbf{Sy})$, the 2-form gauge field $B_{\mu\nu}(\Lambda_2)$, and the dilaton $\Phi(\Lambda_0)$. The massless particles in the R-R sectors are either gauge forms of *even* degree ≤ 4 for Type IIB or gauge forms of *odd* degree < 4 for type IIA. The case of 4-forms is subtle (since they involve a self-duality constraint) and will be discussed at length from various points of view in this book. In Sect. 3.7.3 we shall see its meaning in terms of R-R physical vertices.

The sectors $(\text{NS}+, \text{R}\pm)$ contain a gravitino ($\mathbf{56}_{c,s}$) and a dilatino ($\mathbf{8}_{s,c}$) of *opposite* chiralities. Indeed, in the light-cone quantization the dilatino state has the form

$$\psi_{-1}^i |p; \tilde{\alpha}\rangle (\gamma_i)^{\dot{\alpha}\tilde{\alpha}}, \quad (3.41)$$

where $(\gamma_i)^{\dot{\alpha}\tilde{\alpha}}$ are the $SO(8)$ Dirac matrices. Massless gravitini imply SUSY: when sectors $(\text{NS}+, \text{R}\pm)$, $(\text{R}\pm, \text{NS}+)$ are present the model is supersymmetric in space-time.

3.3 Old Covariant Quantization

The *old covariant quantization* (OCQ) [3] is useful to simplify some arguments and also as a bridge between the manifestly unitary light-cone gauge and the modern BRST quantization. While the light-cone quantization is conceptually similar to the quantization of QED in the manifestly unitary but non-covariant Coulomb gauge, the OCQ formalism is modeled on the Gupta–Bleuler quantization of QED [4]. Our treatment is *not* the historical one: it is a modern re-interpretation of OCQ in the light of Polyakov’s path integral quantization.

We shall mostly write only the equations for the left-moving side of the superstring; similar equations hold for the right-movers too. OCQ for the bosonic string is recovered by focusing on the NS sector and forgetting all world-sheet spinors ψ^μ , T_F , β , γ , etc. and their contributions to the shift constant a . We use the subscript B when referring to a quantity which is specific of the bosonic string.

OCQ as a Covariant Ansatz for Physical States

In OCQ one works in the naive Fock space created by the oscillators of the 2d matter fields, X^μ and ψ^μ , ignoring the FP ghosts which are frozen in their “ground state”. That is, implicitly one makes the *Ansatz* for the physical states

$$|\text{phys}\rangle = |\text{matter}\rangle \otimes |\Omega_{\text{ghost}}\rangle, \quad (3.42)$$

where $|\text{matter}\rangle$ is a vector in the matter Fock space while $|\Omega_{\text{ghost}}\rangle$ is a fixed Fermi/Bose sea state for the b, c and β, γ systems

$$|\Omega_{\text{ghost}}\rangle = \begin{cases} |\Omega_{\text{ghost}}\rangle_{\text{B}} \stackrel{\text{def}}{=} |q = +1\rangle_{b,c} & \text{for the B string} \\ |\Omega_{\text{ghost}}\rangle_{\text{NS}} \stackrel{\text{def}}{=} |q = +1\rangle_{b,c} \otimes |q = -1\rangle_{\beta,\gamma} & \text{for the NS sector} \\ |\Omega_{\text{ghost}}\rangle_{\text{R}} \stackrel{\text{def}}{=} |q = +1\rangle_{b,c} \otimes |q = -\frac{1}{2}\rangle_{\beta,\gamma} & \text{for the R sector.} \end{cases} \quad (3.43)$$

We stress that $|\Omega_{\text{ghost}}\rangle$ is never the ghosts' vacuum $|0\rangle_{\text{ghost}}$ in the CFT sense (the state corresponding to 1) rather they are “vacua” in the *naïve* sense of oscillator algebra.¹²

The validity of the Ansatz (3.42) should be justified a posteriori. This leads to a subtlety which is often missed. The dual of the sea $|q\rangle$ is the sea $|-q - Q\rangle$; cf. Sect. 2.5.3. A sea state and its dual have the “same” physical properties. If the Ansatz can be justified with the ghosts frozen in a sea state, it can also be justified with the ghosts in the dual sea. We then have alternative OCQs obtained by the replacements

$$|q = +1\rangle_{b,c} \rightsquigarrow |q = +2\rangle_{b,c}, \quad |q = -\frac{1}{2}\rangle_{\beta,\gamma} \rightsquigarrow |q = -\frac{3}{2}\rangle_{\beta,\gamma}. \quad (3.44)$$

All choices of seas lead to physically equivalent states. The conventional choice (3.43) selects the ghost seas which are annihilated by the anti-ghost zero-modes b_0 and β_0 (besides all positive ghost modes).

OCQ Physical States

In the OCQ formalism, the world-sheet gauge constraints

$$T_B(z) = \tilde{T}_B(\bar{z}) = T_F(z) = \tilde{T}_F(\bar{z}) = 0 \quad (3.45)$$

are imposed in the *weak sense* of matrix elements between *physical states*

$$\begin{aligned} \langle \text{phys} | T_B | \text{phys}' \rangle &= \langle \text{phys} | T_F | \text{phys}' \rangle = 0, \\ \langle \text{phys} | \tilde{T}_B | \text{phys}' \rangle &= \langle \text{phys} | \tilde{T}_F | \text{phys}' \rangle = 0. \end{aligned} \quad (3.46)$$

In analogy with QED *à la* Gupta–Bleuler, there is a natural way to enforce Eq. (3.46) which is consistent with the superVirasoro commutation relations (2.432)–(2.434): in the NS sector one defines the physical states $|\text{phys}\rangle_{\text{NS}}$ through the conditions¹³

$$(L_0 - a_{\text{NS}})|\text{phys}\rangle_{\text{NS}} = 0, \quad L_n|\text{phys}\rangle_{\text{NS}} = G_r|\text{phys}\rangle_{\text{NS}} = 0, \quad n, r > 0, \quad (3.47)$$

where L_n, G_r are the modes of the *matter* SCFT currents $T_B^{\text{matter}}(z), T_F^{\text{matter}}(z)$. In the bosonic string, the first condition is replaced by $(L_0 - a_B)|\text{phys}\rangle_B = 0$ with a different shift a_B and there are no G_r . We stress that in this section L_m, G_r are the superVirasoro generators of the *matter CFT* containing the modes of ψ^μ and ∂X^μ but no ghost.

The shift of L_0, a_{NS} (resp. a_B) may be computed in two ways: either by the modern argument in Eqs. (3.55)–(3.57) below, or by physical unitarity which requires the zero-point energy of the ghosts to cancel that of the longitudinal oscillators, so that the shifts agree with the light-cone ones; cf. Sect. 3.2. Both methods give

$$a_{\text{NS}} = \frac{1}{2}, \quad a_B = 1. \quad (3.48)$$

¹² That is, they *together with their duals* are precisely the states annihilated by all positive modes.

¹³ As always we write the left-moving conditions only. In the closed superstring, we have similar conditions on the right-movers with $L_n \leftrightarrow \tilde{L}_n, G_r \leftrightarrow \tilde{G}_r$.

From (3.47) we get the NS sector mass-shell condition

$$\alpha' p^2 + N - \frac{1}{2} = 0, \quad (\text{open superstring NS}) \quad (3.49)$$

$$\frac{\alpha'}{4} p^2 + N - \frac{1}{2} = 0, \quad (\text{closed superstring NS}) \quad (3.50)$$

where N is the *matter* level operator

$$N \stackrel{\text{def}}{=} \sum_{n>0} \alpha_{-n}^\mu \alpha_{\mu n} + \sum_{r>0} r \psi_{-r}^\mu \psi_{\mu r}. \quad (3.51)$$

For the boson string replace $-1/2$ by -1 in Eqs. (3.49), (3.50) and omit the Fermi oscillators ψ_r^μ in the definition of the level operator N .

In the R-sector $T_F(z)$ has a zero-mode G_0 . The physical conditions then become

$$L_n |\text{phys}\rangle_R = G_r |\text{phys}\rangle_R = 0, \quad n > 0, \quad r \geq 0. \quad (3.52)$$

The mass-shell condition arises from the equation $G_0 |\text{phys}\rangle_R = 0$: on Ramond states with ghosts frozen in $|\Omega_{\text{ghost}}\rangle_R$, one has

$$G_0^2 = \begin{cases} \alpha' p^2 + N & (\text{open superstring}) \\ \frac{\alpha'}{4} p^2 + N & (\text{closed superstring}) \end{cases} \quad (3.53)$$

without any shift, as one checks by any one of the previous methods, or by using world-sheet supersymmetry.

From Eqs. (3.47)–(3.52) together with $L_n^\dagger = L_{-n}$, $G_r^\dagger = G_{-r}$, we see that Eq. (3.46) holds between any pair of physical states for the *matter* energy–momentum tensor $T_B^{\text{mat}}(z) \equiv \sum_n L_n z^{-n-2}$, except for the shift $L_0 \rightsquigarrow L_0 - a$ of its zero-mode. In the old times this shift was seen as due to a “normal order” ambiguity in the definition of N . Nowadays we understand it as the effect of the ghosts vacuum energy which should be properly taken into account in computing the total energy–momentum tensor

$$T_B(z) = T_B^{\text{mat}}(z) + T_B^{\text{ghost}}(z). \quad (3.54)$$

$T_B(z)$ is the actual operator which is set to zero as a constraint in the covariant quantization *à la* Polyakov.¹⁴ For physical states of the form (3.42), we have

$$0 = \langle \text{phys} | T_B(z) | \text{phys}' \rangle = \langle \text{phys} | T_B^{\text{mat}}(z) | \text{phys}' \rangle + \frac{h_{\text{ghost}}}{z^2} \langle \text{phys} | \text{phys}' \rangle \quad (3.55)$$

where h_{ghost} is the Virasoro weight of the primary state $|\Omega_{\text{ghost}}\rangle$:

¹⁴ Indeed, it is the total central charge c^{tot} which vanishes (in critical dimensions) and so only the constraint $T^{\text{tot}} = 0$ is algebraically consistent (albeit in the weak sense).

$$h_{\text{ghost}} = \begin{cases} h_B = \frac{1}{2} 1(1 + (1 - 4)) = -1 \equiv -a_B \\ h_{\text{NS}} = \frac{1}{2} 1(1 + (1 - 4)) - \frac{1}{2}(-1)(-1 - (1 - 3)) = -\frac{1}{2} \equiv -a_{\text{NS}}, \\ h_R = \frac{1}{2} 1(1 + (1 - 4)) - \frac{1}{2}(-\frac{1}{2})(-\frac{1}{2} - (1 - 3)) = -\frac{5}{8} \equiv -\frac{c_{\text{matter}}}{24}, \end{cases} \quad (3.56)$$

so, Eq. (3.55) yields

$$(L_0 - 1)|\text{phys}\rangle_B = (L_0 - \frac{1}{2})|\text{phys}\rangle_{\text{NS}} = G_0|\text{phys}\rangle_R = 0, \quad (3.57)$$

in terms of the matter L_0 and G_0 . Equation (3.57) shows that the shift of L_0 is due to the ghosts being frozen in a state which is *not* the $SL(2, \mathbb{C})$ -invariant vacuum $|0\rangle$.

Note 3.1 In view of the superVirasoro algebra, the conditions (3.47), (3.52) reduce to

$$\begin{aligned} (L_0 - a_B)|\text{phys}\rangle_B &= L_1|\text{phys}\rangle_B = L_2|\text{phys}\rangle_B = 0 \\ (L_0 - a_{\text{NS}})|\text{phys}\rangle_{\text{NS}} &= G_{1/2}|\text{phys}\rangle_{\text{NS}} = G_{3/2}|\text{phys}\rangle_{\text{NS}} = 0, \\ L_1|\text{phys}\rangle_R &= G_0|\text{phys}\rangle_R = 0. \end{aligned} \quad (3.58)$$

Spurious and Null States

From the above construction, we see that the states of the form

$$(L_{-n} + h_{\text{ghost}} \delta_{n,0})|\text{anything}\rangle \quad \text{or} \quad G_{-r}|\text{anything}\rangle \quad n, r \geq 0 \quad (3.59)$$

are orthogonal to *all* physical states. Such states are called *spurious*. The dual to Eq. (3.58) states that an *on-shell spurious state*¹⁵ may always be written in the form

$$G_{-1/2}|\chi_{1/2}\rangle + G_{-3/2}|\chi_{3/2}\rangle, \quad (\text{NS sector}) \quad (3.60)$$

$$G_{-1}|\chi_1\rangle + L_{-1}|\chi'_1\rangle, \quad (\text{R sector}) \quad (3.61)$$

for some states $|\chi_\lambda\rangle$ satisfying $(L_0 + h_{\text{ghost}})|\chi_\lambda\rangle = \lambda|\chi_\lambda\rangle$.

The transition amplitudes between physical states and spurious ones vanish, and hence spurious states are not part of the dynamics of the superstring. It may happen that a state is both physical and spurious. Such a state, being orthogonal to *all* physical states, is in particular orthogonal to *itself*, i.e. it has zero norm. These states are called *null*. Adding null states $|\text{null}_s\rangle$ to the physical states $|\text{phys}_s\rangle$,

$$|\text{phys}_s\rangle \longrightarrow |\text{phys}'_s\rangle \equiv |\text{phys}_s\rangle + |\text{null}_s\rangle, \quad (3.62)$$

does not change any physical amplitude since

$$\langle \text{phys}'_s | \text{phys}'_t \rangle \equiv \langle \text{phys}_s | \text{phys}_t \rangle. \quad (3.63)$$

¹⁵ That is, a spurious state $|\text{spurious}\rangle$ such that $(L_0 - a)|\text{spurious}\rangle = 0$. Some authors *define* the spurious states to be one-shell. A “spurious” state of the form $(L_0 - a)|\text{anything}\rangle$ or $G_0|\text{anything}\rangle$ is never physical, hence never null. So the two definitions are equivalent for the purposes of the argument in Eqs. (3.62)–(3.65).

Therefore two physical states $|\text{phys}\rangle$ and $|\text{phys}'\rangle$ which differ by a null state cannot be distinguished by any observable, and hence should be considered to be physically equivalent $|\text{phys}\rangle \sim |\text{phys}'\rangle$. In other words, the null states are *redundancies* of the formalism a.k.a. *gauge symmetries*. Indeed, we shall see in Sect. 3.3 that the “obvious” gauge symmetries of spacetime physics, like reparametrization invariance, local supersymmetry, or Yang–Mills symmetry, have the form

$$\delta|\text{phys}\rangle = |\text{null}\rangle, \quad (3.64)$$

for appropriate $|\text{phys}\rangle$ and $|\text{null}\rangle$. The actual *physical* Hilbert space $\mathbf{H}$ is then the quotient of the space H_{phys} of states satisfying the physical conditions (3.47), (3.52) with respect to the space of null states H_{null} ,

$$\mathbf{H} = H_{\text{phys}} / H_{\text{null}}. \quad (3.65)$$

H_{phys} is a subspace of the Hilbert space H_{SCFT} of the matter SCFT theory on the world-sheet. The physical-state conditions (3.47), (3.52) as well as the null state conditions (3.59) are expressed in terms of the action of the superVirasoro generators L_m, G_r . Hence the space of null states ($\equiv$ gauge symmetries) depends only on the superVirasoro representation content of the matter Hilbert space H_{SCFT} . For the simple SCFT model (3.1), it depends on the spacetime dimension d (which is $2/3$ of the Virasoro central charge c) and the shift a_{NS} (which is minus the Virasoro weight of the implicit ghost state).

Our experience with gauge QFTs teaches us that gauge symmetries are crucial to guarantee consistent interactions. It is a safe guess that the most interesting theory—and the one with the best chances of having consistent interactions—is the one with the largest possible set of gauge symmetries. In the following exercise, the reader is invited to check that the requirement of maximal gauge symmetry ($\equiv$ most null states) fixes $d_{\text{crit}} = 10$ and $a_{\text{NS}} = 1/2$ or, in the bosonic case, $d_{\text{crit}} = 26$ and $a_{\text{B}} = 1$.

Exercise 3.1 Show that the existence of null states of the form $G_{-1/2}|\chi\rangle$ implies $a_{\text{NS}} = \frac{1}{2}$, while the existence of null states $(G_{-3/2} + 2G_{-1/2}L_{-1})|\chi\rangle$ implies $d = 10$.

We stress that Eq. (3.64) corresponds to an *infinite* tower of gauge symmetries, involving gauge fields of arbitrary high spin and masses. The “obvious” massless gauge symmetries mentioned above (Yang–Mills, local SUSY, spacetime diffeomorphisms, etc.) constitute just the level zero of the infinite tower.

Unitarity and the No-Ghost Theorem

Unitarity requires the Hermitian product on the physical Hilbert space $\mathbf{H}$, induced from the *indefinite*¹⁶ one in H_{SCFT} , to be *positive-definite*. 2d gauge invariance (if non-anomalous) implies the isomorphism

$$\mathbf{H} \simeq \mathbf{H}_{\text{l.c.}} \quad (3.66)$$

¹⁶ The norm in H_{SCFT} is not positive-definite since the matter 2d SCFT is non-unitary because of the negative metric for the time-like field X^0 and ψ^0 .

Since the light-cone Hilbert space $\mathbf{H}_{\text{l.c.}}$ is positive-definite, if we can prove the isomorphism (3.66) we conclude that $\mathbf{H}$ is also positive-definite, a result known as the *no-ghost theorem* [3, 5, 6]. The proof of the no-ghost theorem is long and technical. Here we limit ourselves to sketch the argument omitting all details. The interested reader may find the proof in its full glory in the **Appendix** to this chapter.

Idea of the Proof By definition, the isomorphism (3.66) (if it exists) is not Lorentz covariant. One has to show that, given any physical state $|\text{phys}\rangle$, we can find a Lorentz frame and a null state $|\text{null}\rangle$ such that, with respect to this frame,

$$|\text{phys}\rangle + |\text{null}\rangle = |\text{matter}\rangle \otimes |\Omega_{\text{ghost}}\rangle \quad \text{with} \quad |\text{matter}\rangle \in \mathbf{H}_{\text{trans}}, \quad (3.67)$$

where $\mathbf{H}_{\text{trans}}$ is the space of the purely transverse states. A *transverse state* $|\psi\rangle \in \mathbf{H}_{\text{trans}}$ is a matter state which satisfies the physical conditions (3.47), (3.52) and does not contain any longitudinal oscillator $\alpha_{-n}^-, \psi_{-r}^-$. Therefore it is created out of a “ground state” by acting with $\alpha_{-n}^+, \psi_{-r}^+$ ($n, r > 0$). A state $|\psi\rangle \in \mathbf{H}_{\text{trans}}$ splits as

$$|\psi\rangle = |\psi\rangle_{\text{l.c.}} + |\text{null}\rangle_{\text{trans}} \quad (3.68)$$

where $|\psi\rangle_{\text{l.c.}} \in \mathbf{H}_{\text{l.c.}}$ is a state with no longitudinal oscillator α_{-n}^+ and $|\text{null}\rangle_{\text{trans}}$ is a *transverse* state with some α_{-n}^+ excited. The state $|\text{null}\rangle_{\text{trans}}$ is orthogonal to $|\psi\rangle_{\text{l.c.}}$ and has zero norm. We stress that Eq. (3.67) holds *if and only if* we are in the critical dimension 10 (26 for the bosonic string).

Having established (3.67), the theorem is reduced to show that $\mathbf{H}_{\text{trans}}$ is positive-definite. This can be proven directly or by showing that the isometry of Hilbert spaces

$$\mathbf{H}_{\text{trans}} \rightarrow \mathbf{H}_{\text{l.c.}}, \quad |\psi\rangle \mapsto |\psi\rangle_{\text{l.c.}} \quad (3.69)$$

is an isomorphism. See the **Appendix** for the proof of both claims.

One can reformulate the result in a nicer way: one defines the Hermitian operator

$$E = N - N_{\text{trans}} \quad (3.70)$$

which measures the difference between the OCQ and the transverse level numbers. By definition a physical state $|\text{matter}\rangle$ belongs to $\mathbf{H}_{\text{trans}}$ iff it is an eigenstate of E of eigenvalue zero. The no-ghost theorem is equivalent to saying that an OCQ physical state $|\psi\rangle$ which is an eigenvector of E with non-zero eigenvalue, is necessarily null. Then all elements of the quotient $\mathbf{H}$ have a representative as in the RHS of (3.67) with $|\text{matter}\rangle$ a zero eigenvector of E : $E|\text{matter}\rangle = 0$.

The *no-ghost theorem* (3.66) is the ultimate justification of the Ansatz (3.42): it shows that the OCQ rules do reproduce the correct physical spectrum as known from the light-cone. It also shows that we need to keep only the states with one choice of the frozen ghost seas, the different choices in Eq. (3.44) leading to distinct (but equivalent) copies of the physical Hilbert space $\mathbf{H}$. We conclude that the OCQ is

equivalent to light-cone quantization and hence suffers the same limitations: *physical states which exist only at zero-momentum are outside the reach of OCQ.*

GSO Projection

The above story is independent of the GSO projection. To get a spectrum which is physically consistent in the fully interacting theory, we have to implement the projection.¹⁷ Since L_n (resp. G_r) commute (anticommute) with the chirality operator $(-1)^F$, physical, spurious, and null states may all be chosen to have definite chirality, and hence there is a well-defined (induced) action of $(-1)^F$ on the physical Hilbert space $\mathbf{H}$, Eq. (3.65), which then splits into *even* and *odd* parts

$$\mathbf{H} = \mathbf{H}_+ \oplus \mathbf{H}_-, \quad (-1)^F|_{\mathbf{H}_\pm} = \pm 1. \quad (3.71)$$

The GSO projection $(-1)^F = +1$ keeps only states whose matter part $|\text{matter}\rangle$ is *odd*

$$(-1)^F|\text{matter}\rangle = -|\text{matter}\rangle, \quad (3.72)$$

since the ghosts' "vacuum" $|\Omega_{\text{ghost}}\rangle$ has chirality -1 , which is the natural value from the "unified" $SO(10, 2)$ matter+ghost current algebra of Sect. 3.1. Indeed, the state

$$|q = -1\rangle_{\beta, \gamma} \equiv e^{-\phi}|0\rangle_{\beta, \gamma} \quad (3.73)$$

from the point of view of the current algebra in Sect. 3.1 has the $SO(10, 2)$ weight $\lambda \equiv (0, 0, 0, 0, 0; -1)$, so its chirality is (cf. Eq. (3.11))

$$(-1)^{\iota \cdot \lambda} = -1. \quad (3.74)$$

This accounts for the "reverse" sign (3.72) in the matter GSO projection.

Example: open superstring massless states

For the massless states, it is enough to consider the truncation of the superVirasoro generators G_r to the zero-mode contributions. For, say, the open superstring:

$$G_0 = (2\alpha')^{1/2} p_\mu \psi_0^\mu + \cdots \quad G_{\pm 1/2} = (2\alpha')^{1/2} p_\mu \psi_{\pm 1/2}^\mu + \cdots \quad (3.75)$$

In the NS sector, a $N = \frac{1}{2}$ level state must have the form

$$|e; k\rangle_{\text{NS}} = e_\mu \psi_{-1/2}^\mu |0; k\rangle_{\text{NS}}, \quad (3.76)$$

where $|0; k\rangle_{\text{NS}}$ stands for the NS oscillators' vacuum state at momentum k , that is,

$$|0; k\rangle_{\text{NS}} \stackrel{\text{def}}{=} e^{ik \cdot X(0)} |0\rangle_{\text{matter}} \otimes |\Omega_{\text{ghost}}\rangle_{\text{NS}}. \quad (3.77)$$

The physical conditions are

$$0 = (L_0 - \tfrac{1}{2}) |e; k\rangle_{\text{NS}} = \alpha' k^2 |e; k\rangle_{\text{NS}} \quad 0 = G_{1/2} |e; k\rangle_{\text{NS}} = (2\alpha')^{1/2} e_\mu k^\mu |e; k\rangle_{\text{NS}}, \quad (3.78)$$

¹⁷ The GSO projection is not the only way to get consistency: see Chap. 5 for full details.

while the massless-level null states have the form

$$G_{-1/2}|0; k\rangle_{\text{NS}} = (2\alpha')^{1/2} k_\mu \psi_{-1/2}^\mu |e; k\rangle_{\text{NS}}. \quad (3.79)$$

Hence the physical NS states at the level $N = 1/2$ are described by

$$k^2 = 0, \quad k \cdot e = 0, \quad e_\mu \simeq e_\mu + \rho k_\mu \quad \rho \in \mathbb{R}, \quad (3.80)$$

Thus (as in the bosonic string) the massless level of the open NS superstring is a massless vector of transverse polarization e_μ . The constraint and the equivalence relation have eliminated the unphysical polarizations, leaving eight physical ones (as in light-cone gauge). Taking into account the ghosts' contribution, Eq. (3.74), this state has $(-1)^F = +1$ and is GSO allowed.

In the R-sector the $N = 0$ states at momentum k_μ are

$$|u; k\rangle_{\text{R}} = u^\alpha e^{ik \cdot X(0)} |\alpha\rangle \otimes |\Omega_{\text{ghost}}\rangle_{\text{R}}, \quad (3.81)$$

where $|\alpha\rangle$ is the matter Ramond vacua, transforming as a spinor of $SO(9, 1)$ of (say) chirality $+1$ for the GSO allowed states. The physical condition is

$$0 = G_0 |u; k\rangle_{\text{R}} = (\alpha')^{1/2} k_\mu (\gamma^\mu)_{\dot{\beta}\alpha} u^\alpha |\dot{\beta}; k\rangle_{\text{R}}, \quad (3.82)$$

and there are no null states at this level. We get the massless Dirac equation for the spinor u_α

$$\gamma^\mu k_\mu u = 0. \quad (3.83)$$

On-shell, only half the components are independent, and hence the propagating massless fermionic degrees of freedom are $16/2 = 8$ as expected from the light-cone gauge.

In the closed string, we have two copies of the story in the **Example** and the massless spectrum agrees with Table 3.1 from the light-cone analysis.

The Dirac–Ramond Equation

The on-shell condition on the matter part $|\text{matter}\rangle_{\text{R}}$ of an R-sector state reads

$$G_0 |\text{matter}\rangle_{\text{R}} \equiv \sqrt{\alpha'} \left(\sqrt{2} \psi_0^\mu p^\mu + M \right) |\text{matter}\rangle_{\text{R}} = 0 \quad (3.84)$$

$$\text{where } M \equiv \frac{1}{\sqrt{2\alpha'}} \sum_{n \neq 0} \alpha_{-n}^\mu \psi_{\mu n}. \quad (3.85)$$

The Lorentz current algebra identifies the spacetime Dirac matrices Γ^μ with the Fermi zero-modes $\Gamma^\mu \equiv \sqrt{2} \psi_0^\mu$. Diagonalizing the operator M , Eq. (3.84) takes the form of an infinite tower of Dirac equations

$$(i \not{\partial} + M) |\text{matter}\rangle_{\text{R}} = 0. \quad (3.86)$$

3.4 OCQ: Physical Conditions Versus 2d Superfields

The first Eq. (3.58) just says that a bosonic string physical state $|\text{phys}\rangle_B$ is mapped by the CFT state-operator correspondence into a matter-sector conformal primary field of weight $h = 1$. In the closed string, taking into account the right-movers, a physical state has (matter) weights $(h, \tilde{h}) = (1, 1)$.

In the same way, Eq. (3.47) says that a NS sector physical state $|\text{phys}\rangle_{\text{NS}}$ is mapped by the CFT state-operator correspondence into a SCFT primary superfield

$$\Phi(z, \theta) = \phi_0(z) + \theta \phi_1(z) \quad (3.87)$$

of weight $h = \frac{1}{2}$. Indeed in the NS sector its modes are half-integral

$$\phi_0(z) = \sum_{r \in \mathbb{Z} + \frac{1}{2}} \frac{\phi_{0,r}}{z^{r+1/2}}, \quad (3.88)$$

and the matter part of the physical NS state has the form

$$|\text{matter}\rangle = \phi_{0,-1/2}|0\rangle \equiv |\phi_0\rangle, \quad (3.89)$$

where $|0\rangle$ is the $SL(2, \mathbb{C})$ -vacuum of the matter SCFT. From (2.444) we get

$$L_m |\text{matter}\rangle = [L_m, \phi_{0,-1/2}]|0\rangle = \frac{1}{2}(1-m)\phi_{0,m-\frac{1}{2}}|0\rangle \quad (3.90)$$

$$G_r |\text{matter}\rangle = \{G_r, \phi_{0,-1/2}\}|0\rangle = \phi_{1,r-\frac{1}{2}}|0\rangle. \quad (3.91)$$

From the mode expansion and the regularity of $|0\rangle$, we know that (cf. (2.445))

$$\phi_{0,s}|0\rangle = \phi_{1,m}|0\rangle = 0 \quad \text{for } s \geq \frac{1}{2}, m \geq 0, \quad (3.92)$$

which, in view of Eqs. (3.90) and (3.91), imply the physical conditions

$$(L_m - \frac{1}{2}\delta_{m,0})|\text{matter}\rangle = G_r|\text{matter}\rangle = 0 \quad m \geq 0, r \geq \frac{1}{2}. \quad (3.93)$$

This shows that, as claimed, the condition of $\phi_0(z)$ being the first component of a conformal superfield of weight $h = \frac{1}{2}$ is equivalent to the physical state conditions (3.47) for the corresponding state $|\phi_0\rangle \otimes |\Omega_{\text{ghost}}\rangle_{\text{NS}}$. The same argument shows that the state (3.89) is (the matter part of) a *OCQ null* state iff the superconformal primary $\phi_0(z)$ is also a descendent, i.e. if it is null in the SCFT sense.

For instance in the case of **Example (3.75)–(3.80)**, we have¹⁸

¹⁸ **Convention:** We reserve the notation X^μ for the component field and use the boldface symbol $\mathbf{X}^\mu(z, \theta)$ for the corresponding superfield. Just as in the bosonic case, $\mathbf{X}^\mu(z, \theta)$ is *not* a conformal superfield, while $D\mathbf{X}^\mu$ and $e^{ik \cdot \mathbf{X}}$ are.

$$\phi_0(z) = e_\mu \psi^\mu(z) e^{ik \cdot X(z)}, \quad \text{i.e.} \quad \Phi(z, \theta) = e_\mu D X^\mu(z, \theta) e^{ik \cdot X(z, \theta)}, \quad (3.94)$$

which is primary iff $e \cdot k = 0$ and has weight $h = \frac{1}{2}$ iff $k^2 = 0$. The state is spurious iff it corresponds to the second component of a superfield, i.e. if the associated operator is the *bottom component* of the superderivative of a superfield, and it is *null* if it is both physical and spurious. In the present example, this means that ϕ_0 has the form

$$D(e^{ik \cdot X}) \Big|_{\theta=0} \equiv ik \cdot D X e^{ik \cdot X} \Big|_{\theta=0} = ik_\mu \psi^\mu(z) e^{ik \cdot X(z)}. \quad (3.95)$$

Thus we recover the three physical conditions in Eq. (3.80).

The NS-NS sector of the closed superstring has two copies of the same story, i.e.

$$\begin{aligned} (L_m - \tfrac{1}{2}\delta_{m,0})|\text{phys}\rangle_{\text{NS}} &= (\tilde{L}_m - \tfrac{1}{2}\delta_{m,0})|\text{phys}\rangle_{\text{NS}} = 0 \\ G_r|\text{phys}\rangle_{\text{NS}} &= \tilde{G}_r|\text{phys}\rangle_{\text{NS}} = 0, \quad m \geq 0, \quad r > 0, \end{aligned} \quad (3.96)$$

and hence a NS-NS physical state corresponds to a conformal superfield $V(z, \bar{z}, \theta, \bar{\theta})$ of conformal weights $(h, \tilde{h}) = (\frac{1}{2}, \frac{1}{2})$. Its *superspace* integral

$$\int d^2z d^2\theta V(z, \bar{z}, \theta, \bar{\theta}) \equiv \int d^2z V_{1,1}(z, \bar{z}) \quad (3.97)$$

is a superconformal invariant. For the bosonic string, the conformal invariant is

$$\int d^2z V(z, \bar{z})_B \quad (3.98)$$

with $V(z, \bar{z})_B$ a CFT primary field of weights $(h, \tilde{h}) = (1, 1)$. Such a field transforms as a 2-form under world-sheet reparametrizations, so its integral is invariant and well-defined—the same comment applies to the RHS of Eq. (3.97) since $V_{1,1}(z, \bar{z})$ is also a Virasoro primary of weights $(1, 1)$.

The insertion of a vertex operator of the form (3.97) or (3.98) in an amplitude is equivalent to an infinitesimal perturbation of the matter part of the world-sheet action

$$S \rightarrow S + \epsilon \int d^2z d^2\theta V + O(\epsilon^2), \quad (\text{superstring}) \quad (3.99)$$

$$S_B \rightarrow S_B + \epsilon \int d^2z V_B + O(\epsilon^2) \quad (\text{bosonic string}). \quad (3.100)$$

In the language of 2d QFT, the physical conditions (3.96) are just the statement that (3.97), (3.98) yield a *marginal* deformation, i.e. a deformation which preserves the conformal invariance of the QFT to leading order in ϵ , while in the superstring case the perturbation manifestly respects supersymmetry (being an integral over superspace). Thus the deformation of the world-sheet theory in the direction of a physical NS-NS operator preserves the full 2d superconformal symmetry.

On the other hand, the condition that the NS-NS $(\frac{1}{2}, \frac{1}{2})$ super-primary operator V is *null* is equivalent to the statement¹⁹

$$\int d^2\theta V = \text{total derivative}, \quad (3.101)$$

so that its addition to the action S does not deform the theory at all. Analogously, in the bosonic case V_B is null iff it is a total derivative. These properties may be seen as the characterization of the bosonic, respectively, NS-NS, physical/null operators/states.

Bosonic String: Integrated Versus Non-integrated Vertices

The structure of the bosonic string physical states in OCQ may be understood in terms of the operator-state correspondence in yet another way.

In the discussion of Sect. 1.8, we considered the asymptotic states as the result of linearizing the equations of motion around a target-space background. Since the e.o.m. are obtained by setting the beta-functions to zero, the solutions to their linearized version are the marginal deformations of the action S_B in the path integral,

$$e^{-S_B} \rightarrow e^{-S_B} \left(1 - \epsilon \int d^2z V(z, \bar{z}) + O(\epsilon^2) \right). \quad (3.102)$$

The integration in d^2z guarantees the invariance under reparametrization; in turn this may be seen as the first step in the integration over the moduli space $\mathcal{M}_{g,n}$ of surfaces with punctures in which we integrate over the points where the primitive $(h, \tilde{h}) = (1, 1)$ fields are inserted.

However in the presence of zero-modes for the c ghost, the integral over $\mathcal{M}_{g,n}$ is subtler than that. When the world-sheet is a sphere ($g = 0$), using the $SL(2, \mathbb{C})$ symmetry we may fix three points (say at 0, 1, and ∞), introducing a finite-dimensional Faddeev-Popov determinant which is reproduced by the insertion of the ghosts $c\tilde{c}$ at these 3 points (cf. chap. 1). We end up with an integral over the positions of the other $n - 3$ insertions. Indeed $\dim \mathcal{M}_{0,n} = n - 3$.

We conclude that in the world-sheet correlation functions, we have *two versions* of the physical vertex of a given physical state of the bosonic string, the *integrated* and *non-integrated* vertex, of the respective form

$$\int d^2z V(z, \bar{z}), \quad \text{and} \quad c(z_0)\tilde{c}(\bar{z}_0) V(z_0, \bar{z}_0) \quad (3.103)$$

with $V(z, \bar{z})$ a matter-sector primary operator of weights $(h, \tilde{h}) = (1, 1)$.

The complex moduli space point of view then gives us a more intrinsic (and modern) interpretation of the ghost “vacuum” $|\Omega_{\text{ghost}}\rangle_B$ (and hence of the shift $L_0 \rightarrow L_0 - 1$). Indeed, we know that²⁰

¹⁹ For instance, in the example (3.95) (writing only the left side) $\int d\theta D e^{ik \cdot X} = \partial e^{ik \cdot X}$.

²⁰ Here $\sigma(z)$ is the free scalar which bosonizes the b, c system; see Sect. 2.5.

$$|c\rangle = \lim_{z \rightarrow 0} c(z)|0\rangle = \lim_{z \rightarrow 0} e^{\sigma(z)}|0\rangle \equiv |q = +1\rangle_{b,c} = |\Omega_{\text{ghost}}\rangle_B \quad (3.104)$$

so that the ghosts' frozen "vacuum" in the OCQ is nothing else than the sea state corresponding to the *local* (i.e. non-integrated) vertex of a physical state. It is this version of the vertex operator which, when inserted at the origin $0 \in \mathbb{C}$, produces the asymptotic **[in]** state of the bosonic string via the state/operator correspondence, and hence represents fully the in-coming physical state. Compare with the discussion in Sect. 1.9. Alternatively we may use the dual ghost vacuum

$$c_0|c\rangle \equiv |\partial c\rangle = |q = +2\rangle_{b,c}. \quad (3.105)$$

Various Forms of Physical NS-NS Vertices

We have seen that the ghost fields in the un-integrated form of the physical vertex just reflect the structure of the ghost "vacuum" which is an implicit factor in the physical OCQ states. In the bosonic string this "vacuum" is $c(0)|0\rangle = e^{\sigma(0)}|0\rangle$, where $|0\rangle$ is the ghosts' SL_2 -vacuum (times the corresponding factor from the right-moving side). Hence in the closed bosonic string, the *local* operator which is the CFT correspondent of the full matter+ghosts state is

$$c(z) \tilde{c}(\bar{z}) V(z, \bar{z}), \quad (3.106)$$

with $V(z, \bar{z})$ a primary with $(h, \tilde{h}) = (1, 1)$.

In the closed superstring the NS-NS ghosts' "vacuum" is

$$c(0) \tilde{c}(0) \delta(\gamma(0)) \delta(\tilde{\gamma}(0))|0\rangle \equiv \delta(c(0)) \delta(\tilde{c}(0)) \delta(\gamma(0)) \delta(\tilde{\gamma}(0))|0\rangle, \quad (3.107)$$

and the state-operator correspondence says that the CFT local operator (inserted at one point, say the origin) corresponding to a physical NS-NS state is

$$c \tilde{c} \delta(\gamma) \delta(\tilde{\gamma}) \phi_{0,0}. \quad (3.108)$$

Comparing with (3.99), we see that (in the closed superstring) the integral in d^2z may be traded for the fermionic ghost factor $c \tilde{c}$, while the integral in $d^2\theta$ may be traded for the corresponding bosonic ghost factor $\delta(\gamma) \delta(\tilde{\gamma})$

$$\int d^2z \rightsquigarrow c \tilde{c}, \quad \int d^2\theta \rightsquigarrow \delta(\gamma) \delta(\tilde{\gamma}), \quad (3.109)$$

in perfect analogy with the bosonic string case. Thus we have many ways of writing the vertices: integrated over the full superspace $d^2z d^2\theta$, integrated in d^2z only, or only in $d^2\theta$, or not integrated at all.²¹ This story extends to a more general and deep phenomenon, *picture changing*, discussed in detail in Sect. 3.7.1 below.

²¹ We can even make different choices on the left and on the right.

Ramond Vertices

The story with physical Ramond operators is subtler, since the vertex operators must contain the bosonized ghost fields—see Eq. (3.22)—and hence cannot be reduced to a simple deformation of the matter SCFT. The R-R vertices correspond to marginal couplings of the full matter+ghosts SCFT which mix the two sectors; these couplings preserve the 2d SUSY by virtue of the Dirac–Ramond equation—although now *not* manifestly since they cannot be written as an integral over superspace—so, in principle, the situation is not different from the NS-NS one, but the details are much less obvious because the new interactions couple ghost and matter d.o.f. in a non-trivial way.

The most promising method to study non-trivial RR backgrounds is to use the *Green–Schwarz formulation* of the superstring [7] (whose quantization is less elementary [8, 9] than the one in the NS-R formalism). The Green–Schwarz formalism is outside the scope of this textbook; the interested reader is referred to [10–12].

3.5 BRST Invariance: Generalities

The modern covariant quantization is based on the BRST method [13–16].

Preliminaries on BRST

On a closed oriented world-sheet there are a left- and a right-moving BRST currents, $j(z)_{\text{BRS}}$ and $\tilde{j}(\bar{z})_{\text{BRS}}$, whose zero-modes $Q = (j_{\text{BRS}})_0$ and $\tilde{Q} = (\tilde{j}_{\text{BRS}})_0$ are the two BRST supercharges. In the presence of boundaries, whose b.c. couple left- and right-movers, only the total BRST charge

$$Q_{\text{BRST}} = Q - \tilde{Q} \quad (3.110)$$

is globally defined. So, conceptually, there is *one* BRST charge Q_{BRST} . However, when working in the perturbative sector of the closed oriented string, the two terms in Eq. (3.110), Q and $\tilde{Q}$, carry different left/right ghost numbers, are separately nilpotent and anticommute, hence may be seen as a pair of independent BRST charges acting separately on the left- and right-moving sectors. This is our viewpoint.²² Open strings are reduced to one sector of this story by the *doubling trick*, as always.

The BRST currents are constructed following the standard rules in terms of the gauge generators, gauge constraints, gauge-fixing, and ghost/anti-ghost fields. As always, the BRST charges $Q, \tilde{Q}$ are *Grassmann-odd* and *Hermitian*. Moreover they anticommute (since left and right degrees of freedom are independent)

$$Q\tilde{Q} + \tilde{Q}Q = 0. \quad (3.111)$$

We shall check momentarily that they are also nilpotent,

²² In math language, we see the BRST cohomology as the cohomology of a *double complex*; the usual BRST cohomology with one BRST charge is its *associated total complex* [17].

$$Q^2 = \tilde{Q}^2 = 0, \quad (3.112)$$

precisely iff we are in the critical dimension, $d = 26$ or $d = 10$ for the bosonic resp. fermionic string. In fact, since the matter sector enters in the BRST supercurrent $j(z)_{\text{BRS}}$ only through its (super)Virasoro currents, $T_B(z)$ and $T_F(z)$, the only aspect that enters in the proof of (3.112) is that the matter system is a (S)CFT with

$$c_{\text{matter}} = 26 \text{ (bosonic string)} \quad \hat{c}_{\text{matter}} = 10 \text{ (superstring)}. \quad (3.113)$$

In the closed oriented string, we are even allowed to have two different (S)CFTs on the left and on the right, since the two BRST currents are totally independent. Equation (3.113) is just the condition that the *total* matter + ghosts (super)Virasoro algebra with generators

$$L_m^{\text{tot}} = L_m^{\text{matter}} + L_m^{\text{ghost}}, \quad G_r^{\text{tot}} = G_r^{\text{matter}} + G_r^{\text{ghost}} \quad (3.114)$$

has zero central charge, cf. (2.484), i.e. that the Weyl anomaly cancels.

To simplify the formulae, we focus on the left-movers; the same story applies to the right-movers. We write $\mathfrak{H}$ for the formal “Hilbert” space²³ of the *totality* of left-moving d.o.f., matter and ghosts. *Roughly speaking*, physical states correspond to the BRST cohomology with coefficient in this total matter-ghost “Hilbert” space. In other words, a (left-moving) state $|\text{phys}\rangle \in \mathfrak{H}$ is physical iff it is *Q-closed*, i.e.

$$Q|\text{phys}\rangle = 0. \quad (3.115)$$

In particular, all *Q-exact* states

$$Q|\text{anything}\rangle \quad (3.116)$$

are physical. Since Q is Hermitian, *Q-exact* states are *orthogonal to all physical states*, hence they have *zero norm*. Therefore two (left-moving) physical states should be identified if they differ by a *Q-exact* state, i.e.

$$|\text{phys}\rangle \sim |\text{phys}\rangle + Q|\text{anything}\rangle. \quad (3.117)$$

Two Subtle Points Often Missed

We used the cautious words “*roughly speaking*” because of two subtleties.

First we have to pay some care to the appropriate space $\mathfrak{V}$ of legitimate vectors $|\text{anything}\rangle$ in the equivalence relation (3.117). If we take $\mathfrak{V}$ too big, we end up with few “physical” states (or none at all); if we take $\mathfrak{V}$ too small, we end up with more “physical” states than actually present in the spectrum. For instance if in the *bosonic* string we take $\mathfrak{V}$ to be the naive space—states with bounded wave-functions in spacetime—we end up with twice as many “physical” states. The reason beyond

²³ We write “Hilbert” between quotes because the inner product in $\mathfrak{H}$ is not positive-definite, so $\mathfrak{H}$ is merely a *pseudo*-Hilbert space.

this doubling is obvious in view of the discussion around Eq. (3.44): as we shall see below, all Q -closed momentum eigenstates with $p_\mu \neq 0$ can be written in the form

$$|\text{transverse}_1\rangle \otimes |q = +1\rangle_{b,c} + |\text{transverse}_2\rangle \otimes |q = +2\rangle_{b,c} + Q|\text{something}\rangle \quad (3.118)$$

for some transverse states $|\text{transverse}_1\rangle$, $|\text{transverse}_2\rangle$, and $|\text{something}\rangle$ a state of bounded wave-function. If it were not for the second term, one would conclude that the BRST cohomology space is isomorphic to the OCQ Hilbert space $\mathbf{H}$ with the ghosts frozen in the standard sea. The second term gives the second copy of the same Hilbert space with the ghosts in the *dual* sea. Since the standard and dual seas have the “same” algebraic properties, naive BRST cohomology cannot distinguish between them, and we get one copy of the OCQ Hilbert space *per choice* of ghosts seas in (3.44). This is double-counting of states. However this doubling problem is solved if we take for $\mathfrak{V}$ the proper larger space (not closed under duality): the actual BRST cohomology consists of a *single copy* of the physical Hilbert space; see [18, 19].

The traditional strategy, however, is to compute the BRST cohomology in the naive space while supplementing the BRST conditions (3.115) with the so-called *subsidiary conditions*

$$b_0|\text{phys}\rangle = \tilde{b}_0|\text{phys}\rangle = \beta_0|\text{phys}\rangle = \tilde{\beta}_0|\text{phys}\rangle = 0. \quad (3.119)$$

These conditions have the effect of selecting the OCQ conventional ghost sea; cf. Eq. (3.44). However this *very ad hoc* procedure is justified only by comparison with OCQ, and hence suffers the same limitations: *naive cohomology with subsidiary conditions yields the correct physical answer only at non-zero-momentum*.

The second subtlety applies only to the superstring; it is related to the picture level phenomenon already mentioned at the end of Sect. 2.5.7. As discussed there, we expect each physical state to appear *infinitely many times* with picture charge

$$q = q_0 + \mathbb{Z}. \quad (3.120)$$

Although this infinite repetition is crucial for having a formalism which allows (in principle) to compute any perturbative superstring process, it is not a signal that the physical spectrum consists of infinitely many copies for the following reason (in addition to the ones which we shall describe below along the lines of Sect. 2.5.7). Each picture q corresponds to an inequivalent representation of the ghost operator algebra; the representation space associated with a given picture q is the full Fock space $\mathcal{F}_q$ obtained by acting with all matter + ghost oscillators on the Bose sea $|q\rangle$. Each $\mathcal{F}_q$ contains a representation of the full algebra of quantum operators (in particular, all observables), and hence the full physics, including the full physical spectrum. Considering all such $\mathcal{F}_q$ (instead of just a single one) is a matter of computational convenience (and formal elegance) not of completeness of the physical description.

BRST Cohomology on Operators

By the CFT state-operator correspondence, the BRST cohomology with coefficients in the state space $\mathfrak{H}$ is equivalent to the BRST cohomology in the space of quantum operators. An operator O is *physical* iff it is Q - and $\tilde{Q}$ -closed

$$[Q, O] = [\tilde{Q}, O] = 0, \quad (3.121)$$

where $[\cdot, \cdot]$ is the commutator/anticommutator depending on the Grassmann parity of O . Two physical operators are *equivalent* iff they differ by BRST-exact operators

$$O \sim O + [Q, \mathcal{A}] + [\tilde{Q}, \tilde{\mathcal{A}}]. \quad (3.122)$$

Then $|O\rangle \equiv O(0)|0\rangle$ is physical as a state iff $O(z)$ is physical as a local operator. More generally, a physical operator maps a physical state into a physical state, and the cohomology class of the resulting state depends only on the cohomology classes of the operator and of the state on which it acts

$$Q O|\text{phys}\rangle \equiv [Q, O]|\text{phys}\rangle = 0 \quad (3.123)$$

$$(O + [Q, \mathcal{A}])(|\text{phys}\rangle + Q|\text{any}\rangle) = O|\text{phys}\rangle + Q|\Psi\rangle \quad (3.124)$$

$$\text{where } |\Psi\rangle = \mathcal{A}(|\text{phys}\rangle + Q|\text{any}\rangle) \pm O|\text{any}\rangle. \quad (3.125)$$

Then any correlation function of physical operators between physical states

$$\langle \text{phys} | O_1 O_2 \dots O_s | \text{phys}' \rangle \quad (3.126)$$

depends only on the BRST classes of the various operators and states involved. In the superstring case (at tree level) we represent such amplitudes as correlations on the sphere with $s + 2$ insertions, where the two insertions, at $z = 0$ and $z = \infty$, produce the in- and out- physical states $|\text{phys}'\rangle$ and $\langle \text{phys}|$ according to radial quantization. We write the action of Q on an operator inserted at z as a contour integral of the current $j(w)_{\text{BRS}}$ along a small contour C_z encircling z

$$[Q, O(z)] = \oint_{C_z} \frac{dw}{2\pi i w} j(w)_{\text{BRS}} O(z). \quad (3.127)$$

Equations (3.123)–(3.126) then follow from standard contour gymnastics. In particular, this shows that the $SL(2, \mathbb{C})$ -invariant vacuum is BRST-invariant, hence physical.

Note 3.2 The $SL(2, \mathbb{C})$ -invariant vacuum $|0\rangle$ is the first example of a BRST-invariant state which does not correspond to an OCQ physical state. This is no contradiction since $|0\rangle$ has zero-momentum. The OCQ description is really incomplete at $p_\mu = 0$.

Note 3.3 We shall discuss BRST invariance of the superstring *before* imposing the GSO projection (which is required in order to have consistent interactions). The

GSO operator $(-1)^F$ commutes with Q_{BRST} , and hence the Q_{BRST} cohomology classes have definite GSO parity $(-1)^F$, and the GSO projection of the cohomology coincides with the cohomology of the GSO projection.

Computing BRST Cohomology: BRST Homotopies

Let Q be any BRST charge, that is, any Grassmann-odd operator such that

$$Q^\dagger = Q \quad \text{and} \quad Q^2 = 0. \quad (3.128)$$

A bosonic operator K (acting on the formal “Hilbert space” $\mathfrak{H}$) is a *BRST homotopy* iff it can be written in the form

$$K = \{Q, R\} \quad (3.129)$$

for some fermionic operator R , that is, if the bosonic operator K is BRST-exact, hence BRST-closed

$$[Q, K] = 0. \quad (3.130)$$

Equation (3.130) shows that we may compute the BRST cohomology in each K -eigenspace

$$\mathfrak{H}_k \equiv \{|\psi\rangle \in \mathfrak{H} : K|\psi\rangle = k|\psi\rangle\}. \quad (3.131)$$

Lemma 3.1 *The BRST cohomology is trivial in all non-zero K -eigenspaces.*

Proof Let $|\psi\rangle \in \mathfrak{H}_{k \neq 0}$ be Q -closed; then

$$|\psi\rangle = \frac{1}{k} K|\psi\rangle = \frac{1}{k} (QR + RQ)|\psi\rangle = Q\left(\frac{1}{k} R|\psi\rangle\right). \quad (3.132)$$

□

The technique to compute the Q -cohomology is to construct a large family of commuting homotopy operators $\{K_a\}$ and look for the cohomology in their common zero eigenspace

$$\mathfrak{H}_* = \{K_a|\psi\rangle = 0 : \forall a\} \subset \mathfrak{H}. \quad (3.133)$$

By the **Lemma** all cohomology classes have a representative in the small space $\mathfrak{H}_*$.

3.6 BRST Quantization of the Bosonic String

We focus on the left side of the closed oriented world-sheet; in particular Q is the left-movers’ BRST charge. Of course we have a corresponding story on the right side. Moreover, when writing explicit formulae, for definiteness we assume that the matter part of the world-sheet theory is just the free bosonic theory

$$\frac{1}{4\pi} \int \partial X^\mu \bar{\partial} X_\mu. \quad (3.134)$$

The discussion is however independent of this assumption, and applies to *all* 2d CFTs with the correct central charge.

The BRST Charge Q We have

$$Q = \oint \frac{dz}{2\pi i} j(z)_{\text{BRS}} \quad (3.135)$$

where in the bosonic string

$$j(z)_{\text{BRS}} = c T(z)^{\text{matter}} + \frac{1}{2} c T(z)^{\text{ghosts}} + \frac{3}{2} \partial^2 c(z). \quad (3.136)$$

The first two terms yield the canonical expression for the BRST charge valid for any gauge system; the last term—which does not contribute to the BRST charge Q —is the “improvement term” added to make the BRST current $j(z)_{\text{BRS}}$ a conformal primary operator of weight 1. The BRST charge carries b, c ghost number 1. In terms of modes

$$Q = \sum_{m \in \mathbb{Z}} \left(c_{-m} L_m^{\text{matter}} - \frac{1}{2} \sum_{n \in \mathbb{Z}} (m-n) : c_{-m} c_{-n} b_{m+n} : \right). \quad (3.137)$$

By general BRST theory, we know that the action of Q on the matter fields is a gauge transformation whose parameter is the ghost field $c(z)$, while the action on the anti-ghost $b(z)$ is the gauge constraint itself; thus

$$[Q, X^\mu(z)] = c \partial X^\mu(z) \quad (3.138)$$

$$\{Q, c(z)\} = c \partial c(z) \quad (3.139)$$

$$\{Q, b(z)\} = T(z), \quad (3.140)$$

where $T(z) = T(z)^{\text{matter}} + T(z)^{\text{ghost}}$ is the total energy–momentum tensor. We write L_m for the modes of the total tensor $T(z)$. The Jacobi identity yields

$$\begin{aligned} [L_m, L_n] &= [L_m, \{Q, b_n\}] = -\{Q, [b_n, L_m]\} + \{b_n, [L_m, Q]\} = \\ &= \{Q, [L_m, b_n]\} + \{b_n, [\{Q, b_m\}, Q]\} = \\ &= (m-n)\{Q, b_{m+n}\} - \{[Q^2, b_m], b_n\} = \\ &= (m-n)L_{m+n} - \{[Q^2, b_m], b_n\}. \end{aligned} \quad (3.141)$$

Comparing with the Virasoro algebra

$$\{[Q^2, b_m], b_n\} = -\frac{c^{\text{tot}}}{12} (m^3 - m) \delta_{m+n} \quad (3.142)$$

$$\Rightarrow [Q^2, b_m] = -\frac{c^{\text{tot}}}{12} (m^3 - m) c_m \quad (3.143)$$

where

$$c^{\text{tot}} = c^{\text{matter}} + c^{\text{ghost}} = d - 26. \quad (3.144)$$

Equation (3.143) shows that Q^2 acts as zero on all local fields if and only if $c^{\text{tot}} = 0$, that is, iff we are in the critical dimension $d = 26$ or, more generally, *iff the matter part of the world-sheet theory is a CFT with $c^{\text{matter}} = 26$* . From now on we assume to be in critical dimension, so that $Q^2 = 0$. Moreover, for technical simplicity, we study the *naive* version of the cohomology with the subsidiary condition

$$b_0|\psi\rangle = 0. \quad (3.145)$$

The bosonic string **no-ghost theorem** [20] may be stated as follows.

Theorem 3.2 *In the bosonic string, the BRST cohomology space at non-zero space-time momentum, with the subsidiary condition (3.145), is isomorphic to the OCQ physical space $\mathbf{H} = H_{\text{phys}}/H_{\text{null}}$. Since the last space is positive-definite, the BRST cohomology space is also positive-definite.*

Proof Since the BRST charge commutes with the spacetime momentum and the BRST quantization is manifestly covariant, in studying the BRST cohomology we may choose the Lorentz frame so that $p^+ = 1$ and $p^i = 0$ (here we use the assumption $p_\mu \neq 0$). We have to show that all BRST cohomology classes satisfying the subsidiary condition have a representative of the form

$$\begin{aligned} & |\text{matter}\rangle \otimes |\Omega_{\text{ghost}}\rangle_B, \quad \text{where } |\Omega_{\text{ghost}}\rangle_B \equiv |c\rangle \equiv c_1|0\rangle \\ & \text{and } (\alpha_n^+ - \delta_{n,0})|\text{matter}\rangle = (L_n - \delta_{n,0})|\text{matter}\rangle = 0 \text{ for } n \geq 0, \end{aligned} \quad (3.146)$$

that is, a state in which the ghosts are frozen in the $q = +1$ b, c Fermi sea while $|\text{matter}\rangle \in \mathbf{H}_{\text{trans}}$. In Eq. (3.146) we have set $L_m \equiv L_m^{\text{matter}}$ for short.

A basis of the Hilbert subspace with the chosen value of spacetime momentum is²⁴

$$\left(\prod_{m>0} L_{-m}^{a_m} \right) \left(\prod_{n>0} (\alpha_{-n}^+)^{b_n} \right) \left(\prod_{k>0} b_{-k}^{c_k} c_{-k}^{d_k} \right) |\psi\rangle \quad (3.147)$$

where $|\psi\rangle$ goes through a basis of states satisfying (3.146), $a_m, b_n \in \mathbb{N}$, and $c_k, d_k = 0, 1$. We write E, N_g for the Hermitian operators which acting on the basis (3.147) have eigenvalues

$$E = \sum_{n>0} n(a_n + b_n), \quad N_g = \sum_{k>0} k(c_k + d_k). \quad (3.148)$$

The operators E, N_g are non-negative. E is an operator in the matter CFT while N_g in the ghosts' CFT. We have to show that all BRST classes have a representative which is a zero eigenvector of E and N_g . It is easy to check that

$$[E, L_n] = -n L_n \quad (3.149)$$

so E maps OCQ physical states into physical states. Then, on general grounds, we must have

$$E = \sum_{n>0} (D_{-n} L_n + L_{-n} D_n) + (D_0 + \text{const})(L_0 - 1) \quad (3.150)$$

²⁴ Here we use the Poincaré–Birkhoff–Witt theorem [21].

for certain operators D_n (with $D_{-n} = D_n^\dagger$) which, by conformal symmetry, should be the modes of a matter CFT holomorphic vector field

$$D(z) = \sum_{n \in \mathbb{Z}} \frac{D_n}{z^{n-1}}. \quad (3.151)$$

This entails that

$$[L_m, D_n] = -(2m + n)D_{m+n}. \quad (3.152)$$

The explicit form of $D(z)$ is known (and intricate); we do not need it, except for the information that the “normal order” constant in Eq. (3.150) is $+1$. This fact can be inferred from the proof of the **Lemma** below, so it can also be fixed without entering in the details of $D(z)$. We stress that the operator $D(z)$ makes sense only when acting on states with $(\alpha_0^+ - 1)|\psi\rangle = 0$ so its definition is Lorentz frame dependent.

We define the fermionic $h = 1$ current $s(z)$ as²⁵

$$s(z) = D(z) b(z), \quad (3.153)$$

and the corresponding *odd* charge

$$s = \oint \frac{dz}{2\pi i} s(z). \quad (3.154)$$

Lemma 3.2 $E + N_g$ is a BRST homotopy. More precisely one has

$$E + N_g = \{Q, s + b_0\}. \quad (3.155)$$

Proof An ugly but straightforward computation; see **BOX 3.2**. □

Let us complete the proof of the **theorem**. From **Lemmas 3.1** and **3.2**, we see that all cohomology classes have a representative which is a zero eigenvector of $E + N_g$; since both operators are non-negative, this means a simultaneous zero eigenvector of E and N_g . This shows that all BRST classes satisfying the subsidiary conditions have a representative of the form (3.146).

On the other hand, all states (3.146) are BRST closed

$$Q(|\text{matter}\rangle \otimes |c\rangle_{b,c}) = \sum_{n \geq 0} (L_n - \delta_{n,0}) |\text{matter}\rangle \otimes c_{-n} |c\rangle_{b,c} = 0. \quad (3.156)$$

Finally we have to show that no state of the form (3.146) is Q -exact. A Q -exact state has zero norm; but the states in (3.146) form the physical Hilbert space $\mathcal{H}$ of OCQ which is positive-definite by the OCQ no-ghost theorem. Hence it does not contain any non-zero Q -exact state. □

BRST Cohomology on Operators

Using the operator/state correspondence and the above **theorem** and the discussion in Sect. 3.4, we find that the closed string BRST closed operators are of two forms

$$c \tilde{c} V(z, \bar{z}), \quad \int_{\Sigma} d^2 z V(z, \bar{z}), \quad (3.157)$$

²⁵ $D(z)$ has the same conformal properties as the ghost $c(z)$ except that it is bosonic instead of fermionic. Then $D(z) b(z)$ has the same conformal properties as the current $c(z) b(z)$.

BOX 3.2 - Proof of Lemma 3.2

In this box the notation $\vdots$ stands for the *naïve* (i.e. oscillator) normal product. In the bosonic string we have

$$\begin{aligned} Q &= \sum_m c_{-m} L_m + \frac{1}{2} \sum_{m,n} (m-n) \vdots b_{-m-n} c_m c_n \vdots - c_0 \\ s &= \sum_n b_{-n} D_n \quad L_m : \text{matter Virasoro generators.} \end{aligned}$$

One has

$$\begin{aligned} \sum_{m,n} \{c_{-m} L_m, b_{-n} D_n\} &= \sum_{m,n} (c_{-m} b_{-n} [L_m, D_n] + \{b_{-n}, c_{-m}\} D_n L_m) = \\ &= \sum_{m,n} (- (2m+n) c_{-m} b_{-n} D_{m+n} + \delta_{m+n,0} D_n L_m) = D_0 L_0 + \\ &+ \sum_{m>0} (D_{-m} L_m + L_{-m} D_m) - \sum_{m>0} [L_{-m}, D_m] - \sum_{m,n} (2m+n) c_{-m} b_{-n} D_{m+n} = \\ &= D_0 L_0 + \sum_{m>0} (D_{-m} L_m + L_{-m} D_m) - \sum_{m,n} (2m+n) \vdots c_{-m} b_{-n} \vdots D_{m+n}, \end{aligned}$$

while

$$\sum_{\ell, m, n} (m-n) \{ \vdots b_{-m-n} c_m c_n \vdots, b_{-\ell} D_\ell \} = 2 \sum_{m,n} (m-n) \vdots b_{-m-n} c_m \vdots D_n$$

Thus

$$\begin{aligned} \{Q, s + b_0\} &= D_0(L_0 - 1) + \sum_{m>0} (D_{-m} L_m + L_{-m} D_m) + L_0 = \\ &= ((D_0 + 1)(L_0 - 1) + \sum_{m>0} (D_{-m} L_m + L_{-m} D_m)) + N_g \equiv E + N_g, \end{aligned}$$

where we used $L_0 = L_0 + N_g - 1$ (the minus 1 being $\hbar$ for the reference Fermi sea).

where $V(z, \bar{z})$ is a $(h, \tilde{h}) = (1, 1)$ conformal primary of the matter CFT. The operator is Q -exact iff V is also a descendent (i.e. null in the Virasoro sense).

In the open string, we have only one side of the story and the BRST closed operators are

$$c V(x), \quad \int_{\partial \Sigma} dx V(x) \quad (3.158)$$

where $x \in \partial \Sigma$ according to the doubling trick. We stress again that the vertices of the physical states are the operators representing non-trivial BRST cohomology classes.

3.7 BRST Quantization of the Superstring

The BRST current $j(z)_{\text{BRS}}$ is

$$j(z)_{\text{BRS}} = c \left(T^{\text{matter}} + \frac{1}{2} T^{\text{ghost}} \right) - \gamma \left(T_F^{\text{matter}} + \frac{1}{2} T_F^{\text{ghost}} \right). \quad (3.159)$$

It is convenient to decompose the BRST charge Q into components Q_g of definite β, γ $U(1)$ charge g

$$Q \equiv \oint \frac{dz}{2\pi i} j(z)_{\text{BRS}} = Q_0 + Q_1 + Q_2, \quad (3.160)$$

where

$$Q_0 = \oint \frac{dz}{2\pi i} (c T_B(X, \psi, \beta, \gamma) + c(\partial c)b) \quad (3.161)$$

$$Q_1 = - \oint \frac{dz}{2\pi i} \frac{1}{2} \gamma \psi^\mu \partial X_\mu \quad (3.162)$$

$$Q_2 = - \oint \frac{dz}{2\pi i} \frac{1}{4} \gamma^2 b. \quad (3.163)$$

Here $T_B(X, \psi, \beta, \gamma)$ is the sum of the matter and β, γ energy–momentum tensors, so that Q_0 is the would-be “bosonic string” BRST charge in which we consider the β, γ system as part of the “matter” CFT. Q_1 is the matter supersymmetry generator with commuting parameter the ghost γ . Q_2 is the term arising from the ghost supercurrent.

Working as in the bosonic case, we arrive at the following commutation relations

$$[Q, X^\mu(z)] = c \partial X^\mu(z) + \frac{i}{2} \gamma \psi^\mu(z) \quad (3.164)$$

$$\{Q, \psi^\mu(z)\} = \left(\frac{1}{2} \partial c \psi^\mu(z) + c \partial \psi^\mu(z) \right) - \frac{i}{2} \gamma \partial X^\mu(z) \quad (3.165)$$

$$\{Q, c(z)\} = c \partial c(z) - \frac{1}{4} \gamma^2(z) \quad (3.166)$$

$$[Q, \gamma(z)] = -\frac{1}{2} \partial c \gamma(z) + c \partial \gamma(z) \quad (3.167)$$

$$\{Q, b(z)\} = T^{\text{tot}}(z) \quad (3.168)$$

$$[Q, \beta(z)] = -T_F^{\text{tot}}(z) \quad (3.169)$$

$$[Q, T^{\text{tot}}(z)] = \frac{1}{8} (d - 10) \partial^3 c \quad (3.170)$$

$$\{Q, T_F^{\text{tot}}(z)\} = -\frac{1}{8} (d - 10) \partial^2 \gamma, \quad (3.171)$$

where, as always, d stands for $\hat{c}$ of the matter SCFT. The action of Q^2 on the matter fields and the FP ghosts c , γ vanishes by construction. Nilpotency of the BRST charges requires that the action of Q^2 on the anti-ghosts b , β also vanishes

$$[Q^2, b(z)] \stackrel{?}{=} 0, \quad [Q^2, \beta(z)] \stackrel{?}{=} 0. \quad (3.172)$$

We see from Eqs. (3.168)–(3.171) that this happens precisely when $\hat{c} = 10$, i.e. when the superstring is critical. For our purposes the most relevant equations are (3.168)–(3.171). They say that *the following relations hold for all $n \in \mathbb{Z}$ and $r \in \mathbb{Z} + \nu$ in a critical superstring*

$$\{Q, b_n\} = L_n \quad [Q, \beta_r] = G_r \quad (3.173)$$

$$[L_n, Q] = 0 \quad \{G_r, Q\} = 0, \quad (3.174)$$

where L_n , G_r are the total superVirasoro generators of the matter + ghosts SCFT. From now on we assume the superstring to be *critical*. In terms of modes we have

$$\begin{aligned} Q = & \sum_{m \in \mathbb{Z}} c_{-m} L_m^{\text{matt}} + \sum_{r \in \mathbb{Z} + \nu} \gamma_{-r} G_r^{\text{matt}} - \frac{1}{2} \sum_{m, n} (n - m) : b_{-m-n} c_m c_n : + \\ & + \frac{1}{2} \sum_{m, r} \left((2r - m) : \beta_{-m-r} c_m \gamma_r : - : b_{-m} \gamma_{m-r} \gamma_r : \right) + a c_0, \end{aligned} \quad (3.175)$$

where the constant a depends on the “ordering prescription” $:(\cdots):$ we use, i.e. with respect to which Fermi/Bose seas we define the creation/annihilation modes. $a = 0$ in all sectors if $:(\cdots):$ is defined with respect to the SL_2 vacuum $|0\rangle$.

BRST Invariance of the SL_2 Vacuum

Recall that in a CFT the $SL(2, \mathbb{C})$ invariant vacuum $|0\rangle$ is defined by the condition that $\Phi(z)|0\rangle$ is regular as $z \rightarrow 0$ for all fields $\Phi(z)$. For primaries of weight h

$$\sum_{n+h \in \mathbb{Z}} \frac{\Phi_n}{z^{n+h}} |0\rangle \text{ regular at } z = 0 \quad \Leftrightarrow \quad \Phi_n |0\rangle = 0 \text{ for } n \geq 1 - h. \quad (3.176)$$

Hence

$$b_n |0\rangle = 0 \quad n \geq -1 \quad c_n |0\rangle = 0 \quad n \geq 2 \quad (3.177)$$

$$\beta_n |0\rangle = 0 \quad n \geq -\frac{1}{2} \quad \gamma_n |0\rangle = 0 \quad n \geq \frac{3}{2}, \quad (3.178)$$

that is, in terms of Fermi/Bose seas,

$$|0\rangle = |0\rangle_{\text{matter}} \otimes |q = 0\rangle_{b,c} \otimes |q = 0\rangle_{\beta,\gamma}. \quad (3.179)$$

The dual vacuum is the Fermi/Bose sea

$$|0\rangle_{\text{matter}} \otimes |q = 3\rangle_{b,c} \otimes |q = -2\rangle_{\beta,\gamma} \equiv e^{3\sigma-2\phi}|0\rangle, \quad (3.180)$$

where σ is the scalar which bosonizes the b, c system, $c(z) = e^{\sigma(z)}$, $b(z) = e^{-\sigma(z)}$, and ϕ is the scalar which bosonizes the $-\beta\gamma$ $U(1)$ current.

The states in the SCFT Hilbert space H are obtained by acting with the corresponding operators O on the SL_2 vacuum $|0\rangle$: this is the natural vacuum from the viewpoint of the CFT state-operator correspondence. Most authors use other “vacua” which differ in the ghost structure, i.e. they correspond to Fermi/Bose seas of different levels. The choice of the reference Fermi/Bose “vacuum” is just a matter of convention. One passes from one such “vacuum” to another one by multiplying the reference state by the sea-level changing operator $e^{p\sigma+q\phi}$ with the appropriate p, q .

The BRST invariance of the SL_2 vacuum $|0\rangle$ is tautological

$$j_{\text{BRS}}(z)|0\rangle \text{ regular as } z \rightarrow 0 \Rightarrow Q|0\rangle = \oint \frac{dz}{2\pi i} j_{\text{BRS}}(z)|0\rangle = 0. \quad (3.181)$$

By Hermitian conjugation we also have [22]

$$Q e^{3\sigma-2\phi}|0\rangle = 0. \quad (3.182)$$

Exercise 3.2 Prove Eq. (3.182) directly using ghost field OPEs.

The correlation functions

$$\langle 0|e^{3\sigma-2\phi} V_1 \dots V_s|0\rangle \quad (3.183)$$

are BRST-invariant²⁶ provided the inserted vertices V_i are BRST-closed, that is $[Q, V_i] = 0$ and carry *zero ghost charges*. The amplitude (3.183) corresponds to the path integral with operator insertions V_i computed on the sphere, i.e. to the string tree level. In particular, comparing with Sect. 2.5.7, we see that the insertion of the operator

$$e^{3\sigma-2\phi} \rightsquigarrow c \partial c \partial^2 \bar{c} \delta(\gamma) \delta(\partial\gamma) \quad (3.184)$$

has the effect of soaking up the three zero-modes of c and the two zero-modes of γ , producing a finite non-zero physical amplitude.

BRST Cohomology on States

With respect to the case of the bosonic string, the new ingredient of the BRST cohomology in the superstring is the *picture charge*.

We write $\mathfrak{F}(q)$ for the generalized Fock space obtained by acting with the oscillators of the free fields X^μ , ψ^μ , b, c, β , and γ on the state

$$|q\rangle \stackrel{\text{def}}{=} \begin{cases} c(0)e^{q\phi(0)}e^{ip\cdot X(0)}|0\rangle & \text{(NS)} \\ u^\alpha S_\alpha(0)c(0)e^{q\phi(0)}e^{ip\cdot X(0)}|0\rangle & \text{(R)}, \end{cases} \quad (3.185)$$

²⁶ That is, independent of the representative of the V_i in their Q -cohomology class.

that is, the β , γ Bose sea of level q times the b , c Fermi sea of level 1, boosted at momentum $p_\mu \neq 0$. In the Ramond sector, the matter part of $|q\rangle$ is obtained by acting on the SL_2 vacuum with an $SO(10)$ spinor field $u^\alpha S_\alpha$ which implements the Ramond b.c. on the 2d fields ψ^μ . To simplify the notation we leave implicit the dependence of the states on p_μ and u^α . We know from Sect. 3.1 that $q \in \mathbb{Z}$ in the NS sector, and $q \in \mathbb{Z} + \frac{1}{2}$ in the R-sector. We refer to the seas of level $q = -1$ and $q = -\frac{1}{2}$, of the OCQ frozen “vacua”, as the *standard vacua*. We expect that there is a copy of the physical states with all possible picture charges (equal to the Bose sea level q).

“Small” Versus “Large” Space The Fock space $\mathfrak{F}(q)$ is more properly defined as the cyclic module of the algebra $\mathcal{A}_{\text{small}}$ generated by the state $|q\rangle$ (for given p^μ and u^α)

$$\mathfrak{F}(q) \Big|_{\text{fixed}} p^\mu, u^\alpha = \mathcal{A}_{\text{small}} |q\rangle, \quad (3.186)$$

where $\mathcal{A}_{\text{small}}$ is the algebra generated by the operators $\partial X^\mu(z)$, $\psi^\mu(z)$, $b(z)$, $c(z)$, $\beta(z)$, and $\gamma(z)$. It is crucial that $\mathfrak{F}(q)$ is a module of the *small* operator algebra $\mathcal{A}_{\text{small}}$, and not of the *large* one $\mathcal{A}_{\text{large}}$ generated by all operators in the bosonized setup

$$\mathcal{A}_{\text{large}} = \left\langle \partial X^\mu(z), \psi^\mu(z), b(z), c(z), e^{\pm\phi(z)}, \xi(z), \eta(z) \right\rangle. \quad (3.187)$$

We will see below that the Q -cohomology in the large module $\mathcal{A}_{\text{large}}|q\rangle$ is *trivial*, i.e. all Q -closed states are Q -exact. Since the small module is a submodule of the large one, all physical states may be written as Q of something in the large module. In standard BRST quantization, all amplitudes between Q -exact states vanish; then one may suspect that all physical amplitudes—which are just special instances of amplitudes between Q -exact states of the large system—are automatically zero. This is not the case: Q is not Hermitian in the “large” Hilbert space hence not a BRST charge. In the “small” module, Q is Hermitian: it kills both the SL_2 vacuum $|0\rangle$ and its Hermitian conjugate in the small module $e^{3\sigma-2\phi}|0\rangle$; cf. Eq. (3.182). In the large module, the Hermitian conjugate of $|0\rangle$ is $e^{3\sigma-2\phi}\xi_0|0\rangle$ which is not Q -closed.

Computing the Cohomology From (3.174) we see that Q commutes with

$$L_0 = \frac{1}{2}p^2 + N + h_q, \quad (3.188)$$

where p_μ is the spacetime momentum and $N = N_{\text{matter}} + N_{\text{ghost}}$ the *total* level; here

$$N_{\text{matter}} = \sum_{n>0} n \alpha_{-n}^\mu : \alpha_{\mu n} : + \sum_{r>0} r : \psi_{-r}^\mu \psi_{\mu r} : \quad (3.189)$$

$$N_{\text{ghosts}} = \sum_{n>0} n (: b_{-n} c_n : + : c_{-n} b_n :) + \sum_{r>0} r (: \beta_{-r} \gamma_r : - : \gamma_{-r} \beta_r :), \quad (3.190)$$

where the boldface symbol $:\cdots:$ stands for normal order with respect to the standard seas not to be confused with the CFT normal order $\vdots\cdots\vdots$; that is,

$$\begin{cases} N|q = -1\rangle = 0 & \text{(NS)} \\ N|q = -\frac{1}{2}\rangle = 0 & \text{(R)}. \end{cases} \quad (3.191)$$

The constant h_q in (3.188) is the weight of the cyclic state (3.185) equal to the weight h_{ghost} of the standard sea, Eq. (2.256), plus, for the R-sector, the dimension of the matter spin field $S_\alpha(z)$

$$h_{\text{ghost}} = \begin{cases} -\frac{1}{2} & \text{(NS)} \\ -\frac{5}{8} & \text{(R)}, \end{cases} \quad h_q = \begin{cases} -\frac{1}{2} & \text{(NS)} \\ 0 & \text{(R)}. \end{cases} \quad (3.192)$$

By the very notion of normal order, in the standard Fock spaces $\mathfrak{F}(-\frac{1}{2})$ and $\mathfrak{F}(-1)$, the spectrum of N is *non-negative*: it is $\mathbb{N}$ for the R-sector and $\frac{1}{2}\mathbb{N}$ for the NS one. The states with $N = 0$ are obtained by acting on the cyclic state $|q\rangle$ with the zero-modes of the matter and ghost fields.

Since Q commutes with p_μ and L_0 , from Eq. (3.188) we see that it commutes with p^2 and N separately. Q also commutes with the Fermi parity $(-1)^F$ which enforces the GSO projection. The world-sheet parity Ω interchanges Q and $-Q$.

The Charge $E + N_g$ The current $s(z)$, defined for the bosonic string in Eq. (3.153), has a SUSY completion

$$s(z) = D(z)b(z) + \beta(z)B(z), \quad s = \oint \frac{dz}{2\pi i} s(z), \quad (3.193)$$

where $B(z)$ is a fermionic chiral current with $h = -\frac{1}{2}$ [5, 6, 23]; see **Appendix**. Then, as in the bosonic case,

$$E + N_g \equiv \{Q, s + b_0\} \equiv \{Q, s\} + L_0. \quad (3.194)$$

When acting on states satisfying the subsidiary conditions

$$b_0|\psi\rangle = \tilde{b}_0|\psi\rangle = \beta_0|\psi\rangle = \tilde{\beta}_0|\psi\rangle = 0, \quad (3.195)$$

$E + N_g$ is just the difference of the level of all (matter and ghost) oscillators minus the level of the transverse oscillators. That is, as in the bosonic case, the zero eigenvectors of $E + N_g$ satisfying (3.195) are transverse states. The proof of these identities is similar to the one for the bosonic string in Sect. 3.6 but significantly longer; see [5, 6, 23].

3.7.1 Q -Homotopies: Picture Changing

As before we limit ourselves to the left-moving part of the equations. Since $[Q, P_\mu] = 0$ we may study the BRST cohomology at fixed spacetime momentum. The Fock space $\mathfrak{F}(q)$ consists of states of fixed *non-zero* momentum p_μ . We write

$$H(q)_{\text{BRST}} \stackrel{\text{def}}{=} Z(q)_{\text{BRST}} / Q \mathfrak{F}(q) \cap Z(q)_{\text{BRST}} \quad (3.196)$$

$$Z(q)_{\text{BRST}} \stackrel{\text{def}}{=} \left\{ |\psi\rangle \in \mathfrak{F}(q), Q|\psi\rangle = b_0|\psi\rangle = \beta_0|\psi\rangle = 0 \right\} \quad (3.197)$$

for the BRST cohomology space in picture q and momentum $p_\mu \neq 0$ with the subsidiary conditions imposed. We make the following two claims.

Claim 3.1 *At non-zero spacetime momentum $p_\mu \neq 0$, the BRST cohomology in standard pictures is isomorphic to the physical Hilbert space of OCQ*

$$H(-1)_{\text{BRST}} \simeq \mathbf{H}_{\text{NS}}, \quad H(-\tfrac{1}{2})_{\text{BRST}} \simeq \mathbf{H}_R. \quad (3.198)$$

The BRST no-ghost theorem then follows from the OCQ one.²⁷

Claim 3.2 *The BRST cohomology is independent of the picture, up to isomorphism,*

$$H(q)_{\text{BRST}} \simeq H(q+1)_{\text{BRST}}. \quad (3.199)$$

These statements hold both for operators and states by the CFT state/operator isomorphism. We have a copy of the physical states in each picture, as expected.

Note 3.4 We stated (3.198) only when it makes sense, i.e. for states with $p_\mu \neq 0$. All states describing *propagating* degrees of freedom have non-zero-momentum, so by restricting to $p_\mu \neq 0$ we do not miss any d.o.f. which is *local* in spacetime. However we may still have BRST-invariant *non-propagating*²⁸ spacetime d.o.f. which are *invisible* in OCQ and light-cone. As we shall see, BRST-invariant vertices frozen at zero-momentum exist: they represent *quasi-topological* d.o.f. in spacetime which have fundamental physical implications even at the non-perturbative level.

Note 3.5 Ultimately the justification of the subsidiary conditions (3.119) is that computing cohomology in the *naive* space with these conditions yields a cohomology space isomorphic to *one* copy of light-cone Hilbert space. This argument cannot be applied to the $p_\mu = 0$ sector, where we are forced to compute the BRST cohomology in the proper state space without ad hoc extra conditions. However when $p_\mu = 0$ the wave-function is constant (hence bounded) in spacetime, and the proper space coincides with the naive one—but now *without* subsidiary conditions. Indeed imposing these conditions in the $p_\mu = 0$ sector would lead to contradictions [26].

The proof of the two **Claims** is by suitable BRST homotopies. **Claim 3.1** follows from the homotopy (3.194) together with the fact that E and N_g have a non-negative spectrum in the standard picture and the transverse physical states of OCQ satisfy

$$(L_n - \nu \delta_{n,0})|\psi\rangle = \alpha_m^+|\psi\rangle = \psi_r^+|\psi\rangle = 0 \quad n \geq 0, m, r > 0. \quad (3.200)$$

²⁷ For other proofs of the BRST no-ghost theorem, see [24, 25].

²⁸ Non-propagating global degrees of freedom will be called *quasi-topological*.

As in the bosonic case, on-shell transverse states are automatically Q -closed. Indeed, for all states with the ghosts frozen in their “vacuum”

$$|\psi\rangle = |\text{matter}\rangle \otimes |\Omega_{\text{ghost}}\rangle, \quad (3.201)$$

we have (see Eq. (3.192) for the definition of h_{ghost})

$$\begin{aligned} Q|\psi\rangle &= \sum_{n \geq 0} (L_n^{\text{matt}} + h_{\text{ghost}} \delta_{n,0}) |\text{matter}\rangle \otimes c_{-n} |\Omega_{\text{ghost}}\rangle + \\ &+ \sum_{r \geq 0} G_r^{\text{matt}} |\text{matter}\rangle \otimes \gamma_{-r} |\Omega_{\text{ghost}}\rangle. \end{aligned} \quad (3.202)$$

If the state $|\text{matter}\rangle$ is physical in the OCQ sense (in particular, transverse), we get

$$Q|\psi\rangle = 0. \quad (3.203)$$

On the other hand, a non-zero on-shell transverse vector $|\psi\rangle$ cannot be Q -exact: a Q -exact state has zero norm, while the space of transverse states is positive-definite.

Picture Changing

To show **Claim 3.2**, we have to prove the *picture independence* of the Q -cohomology, Eq. (3.199). We recall that we have two distinct operator algebras $\mathcal{A}_{\text{small}}$ and $\mathcal{A}_{\text{large}}$. The “small” one is the actual physical algebra, but computations are best performed in the bigger framework $\mathcal{A}_{\text{large}}$. The essential difference between the two algebras is that the “large” one contains the full free Fermi field $\xi(z)$, a CFT primary of weight *zero*, whereas the “small” algebra contains only its non-zero-modes, i.e. only the 1-form $\partial\xi(z)$. Consider the operator²⁹

$$\begin{aligned} X(z) &\stackrel{\text{def}}{=} 2\{Q, \xi(z)\} = 2c\partial\xi + 2T_F^{\text{matt}} e^\phi + \frac{1}{2} e^{2\phi} b \partial\eta + \frac{1}{2} \partial(e^{2\phi} b\eta) = \\ &= 2T_F^{\text{tot}} e^\phi - \frac{1}{2} (\partial b)\eta e^{2\phi} = 2e^\phi T_F^{\text{matt}} + \text{terms with only ghosts}, \end{aligned} \quad (3.204)$$

where T_F^{matt} (T_F^{tot}) is the matter supercurrent (resp. the total supercurrent). From its explicit expression, we see that $X(z)$ is an operator in the *small* algebra. It is Q -closed by construction. In the large algebra $X(z)$ is also Q -exact, but not in the small algebra, since there is no $\xi(z)$ in that algebra. Thus $X(z)$ represents a non-trivial Q -cohomology class in the physical algebra $\mathcal{A}_{\text{small}}$. Notice however,

$$\partial X(z) = 2\{Q, \partial\xi(z)\} = \text{\textit{\textbf{Q-exact in } } } \mathcal{A}_{\text{small}}, \quad (3.205)$$

²⁹ **Warning:** In this paragraph T_F is normalized to be $T_F/2$ the usual one. This lead to the standard FMS formulae [22] for the picture changing. Other authors have formulae which differ by factors 2.

since $\partial\xi(z)$ belongs to $\mathcal{A}_{\text{small}}$. Thus, if $|\psi_a\rangle$, $a = 1, 2$ are Q -closed states, and the O_i are Q -closed operators, the amplitude

$$\langle\psi_1|X(z)O_1\ldots O_s|\psi_2\rangle \quad \text{is } z \text{ independent.} \quad (3.206)$$

This property extends from the tree-level amplitudes to *all* (connected) amplitudes: we already mentioned that $\xi(z)$, having $h = 0$, has *precisely one zero-mode* on any *connected* world-sheet Σ of whatever topology (compact or *non-compact*, oriented or *non-oriented*).

$X(z)$ has the properties which—in a “sound” 2d CFT—uniquely characterize the identity operator 1: it has zero weight $h = 0$, and is translationally invariant $\partial X \simeq 0$. A 2d experimental physicist has no way to distinguish X from 1. However, X is definitely *not* 1, since it has different quantum numbers: X has picture charge $+1$, not 0. Thus multiplication by X leaves the 2d physics essentially unchanged but increases the picture charge by 1. Therefore X is the natural candidate for the isomorphism between different pictures, Eq. (3.199),

$$|\psi, q\rangle \xrightarrow{\text{picture changing}} \lim_{z \rightarrow 0} X(z)|\psi, q\rangle \equiv |\psi, q+1\rangle. \quad (3.207)$$

The map $X: |\psi, q\rangle \mapsto |\psi, q+1\rangle$ induces a map in Q -cohomology

$$H(q)_{\text{BRST}} \xrightarrow{X} H(q+1)_{\text{BRST}}, \quad (3.208)$$

since $X(z)$ was *defined to be* a Q -homotopy in $\mathcal{A}_{\text{large}}$; cf. Eq. (3.204).

To show that X is an isomorphism in cohomology reduces to construct an inverse picture-changing operator, Y , such that

$$XY = YX = 1 \quad \text{and} \quad [Q, Y] = 0, \quad (3.209)$$

or, more precisely,

$$\lim_{z_1 \rightarrow z_2} X(z_1)Y(z_2) = 1. \quad (3.210)$$

Claim 3.3 $Y(z)$ is given by the formula [27]

$$Y(z) = 2c(z)\partial\xi(z)e^{-2\phi(z)}. \quad (3.211)$$

Note that $Y(z)$ has dimension 0 (recall that $e^{-2\phi}$ has dimension zero by (2.256)), as expected for an operator which should share most of the properties of 1 and has picture charge -1 (the picture charges of $\partial\xi$ and $e^{-2\phi}$ are $+1$ and -2 , respectively).

Proof Let us compute the OPEs of each term in the RHS of (3.204) with $Y(w)$ using bosonization: $c = e^\sigma$, $b = e^{-\sigma}$, $\eta = e^{-\chi}$, and $\xi = e^\chi$. The first two terms in the RHS 1st line of (3.204) give nothing as $z \rightarrow w$

$$c(z)\partial\xi(z)c(w)\partial\xi(w)e^{-2\phi(w)} = O((z-w)) \quad (3.212)$$

$$T_F^{\text{matt}}(z)e^{\phi(z)}c(w)\partial\xi(w)e^{-2\phi(w)} = O((z-w)^2). \quad (3.213)$$

On the other hand, we have

$$\begin{aligned} & \partial_z \left(e^{2\phi(z)} b(z) \eta(z) \right) c(w) \partial \xi(w) e^{-2\phi(w)} = \\ &= -\partial_z \left[(z-w)^4 e^{2(\phi(z)-\phi(w))} \frac{e^{-\sigma(z)+\sigma(w)}}{z-w} \partial_w \left(\frac{e^{-\chi(z)+\chi(w)}}{z-w} \right) \right] = \\ &= -\partial_z \left[(z-w) + O((z-w)^2) \right] = -1 + O(z-w), \end{aligned} \quad (3.214)$$

and, likewise,

$$\begin{aligned} & e^{2\phi(z)} b(z) \partial \eta(z) c(w) \partial \xi(w) e^{-2\phi(w)} = \\ &= - \left[(z-w)^4 \frac{1}{z-w} \partial_z \partial_w \frac{1}{z-w} + \dots \right] = 2 + O(z-w) \end{aligned} \quad (3.215)$$

so

$$\lim_{z \rightarrow w} X(z) Y(w) = 1. \quad (3.216)$$

From the last equation, one infers that $[Q, Y(z)] = 0$ and that $Y(z_1) - Y(z_2)$ is Q -exact

$$Y(z_1) = Y(z_2) + 2 \left\{ Q, Y(z_2) \left(\int_{z_1}^{z_2} \partial \xi(z) \right) Y(z_1) \right\}. \quad (3.217)$$

□

Note 3.6 The picture-changing operation has a more intrinsic interpretation in terms of super-geometry of super-Riemann surfaces; see Sect. 10.2 for a sketch. In that framework, the operations of decreasing and increasing the picture charge with respect to the standard level -1 for NS (resp. $-\frac{1}{2}$ for R) have different meanings. See appendix B of [28] for a discussion.

Changing Picture of Operators: The Rearrangement Lemma

By the state-operator correspondence, picture changing of states can be seen as a change of picture of operators. We write $O_{(q)}$ for the copy of the operator in picture q . For a BRST-invariant operator $O_{(q)}(z)$ we have

$$\begin{aligned} O_{(q+1)}(z) &= \lim_{w \rightarrow z} X(z) O_{(q)}(z) \equiv 2[Q, \xi O_{(q)}(z)] = \\ &= 2 \oint \frac{dw}{2\pi i} j_{\text{BRS}}(w) \xi(z) O_{(q)}(z). \end{aligned} \quad (3.218)$$

A BRST-invariant amplitude has form

$$\langle O_{(q_1)}^{a_1}(z_1) O_{(q_2)}^{a_2}(z_2) \dots O_{(q_s)}^{a_s}(z_s) \rangle_{\Sigma}, \quad (3.219)$$

where the suffixes a_i label the different BRST-closed operators. The total picture charge $\sum_i q_i$ should agree with the net number of β , γ zero-modes, as predicted by the Riemann–Roch theorem

$$\sum_i q_i = -\chi(\Sigma) \equiv 2(g-1). \quad (3.220)$$

The isomorphism (3.199) says that all pictures are physically equivalent, and then the amplitude (3.219) should be *independent* of the arbitrary choice of the q_i (as long as it is consistent with the Riemann–Roch theorem (3.220)), i.e. we must have

$$\langle \cdots \mathcal{O}_{(q_i+1)}^{a_i}(z_i) \cdots \mathcal{O}_{(q_j-1)}^{a_j}(z_j) \cdots \rangle = \langle \cdots \mathcal{O}_{(q_i)}^{a_i}(z_i) \cdots \mathcal{O}_{(q_j)}^{a_j}(z_j) \cdots \rangle \quad (3.221)$$

for all i, j . Equation (3.221) is the *picture rearrangement lemma*. The proof is easy:

$$\begin{aligned} \langle \cdots \mathcal{O}_{(q_i+1)}^{a_i}(z_i) \cdots \mathcal{O}_{(q_j-1)}^{a_j}(z_j) \cdots \rangle &= \langle \cdots X(z_i) \mathcal{O}_{(q_i)}^{a_i}(z_i) \cdots \mathcal{O}_{(q_j-1)}^{a_j}(z_j) \cdots \rangle = \\ &= \langle \cdots X(z_j) \mathcal{O}_{(q_i)}^{a_i}(z_i) \cdots \mathcal{O}_{(q_j-1)}^{a_j}(z_j) \cdots \rangle = \langle \cdots \mathcal{O}_{(q_i)}^{a_i}(z_i) \cdots X(z_j) \mathcal{O}_{(q_j-1)}^{a_j}(z_j) \cdots \rangle \\ &= \langle \cdots \mathcal{O}_{(q_i)}^{a_i}(z_i) \cdots \mathcal{O}_{(q_j)}^{a_j}(z_j) \cdots \rangle, \end{aligned} \quad (3.222)$$

where we used that that $|0\rangle$ is Q -closed and that $X(z_i) - X(z_j)$ is Q -exact, so it vanishes when inserted in a Q -invariant amplitude.

Example: Changing Picture of Standard-Picture NS States

Let $|\psi\rangle$ be a physical NS state in the standard $q = -1$ picture. It satisfies

$$G_r^{\text{matt}}|\psi\rangle = (L_n^{\text{matt}} - \tfrac{1}{2}\delta_{n,0})|\psi\rangle = 0, \quad (3.223)$$

so the matter part of the state is the SCFT state associated with a matter superconformal primary operator $V(z)$ with $h = \frac{1}{2}$; see Sect. 3.4. Taking into account that the ghosts are frozen in the standard-picture Fermi/Bose sea, the physical NS state corresponds to the matter+ghost local operator

$$c V_{(-1)}(z) \stackrel{\text{def}}{=} c e^{-\phi} V(z). \quad (3.224)$$

Let us check that it is BRST-invariant. As in Eqs. (3.161)–(3.163), we write

$$Q = Q_0 + Q_1 + Q_2. \quad (3.225)$$

Q_0 is the world-sheet reparametrization BRST charge, and hence anticommutes with $c V_{(-1)}(z)$ in view of the “identification” in Eq. (3.109).³⁰ We are left with Q_1 and Q_2 whose associate currents we write as $j_1(z)$ and $j_2(z)$, respectively. Now,

³⁰ The skeptical reader is referred to Eqs. (3.239) and (3.240).

$$\begin{aligned}
& \oint \frac{dw}{2\pi i} j_2(w) \xi(z) c(z) V(z) e^{-\phi(z)} = \\
& = -\frac{1}{4} \oint \frac{dw}{2\pi i} \left(b(w) c(z) \right) \left(e^{2[\phi(w)-\chi(w)]} e^{-\phi(z)+\chi(z)} \right) V(z) = \\
& = -\frac{1}{4} \oint \frac{dw}{2\pi i} \left(\frac{1}{w-z} + \dots \right) \left(e^{\phi(z)-\chi(z)} + \dots \right) V(z) = -\frac{1}{4} \gamma(z) V(z),
\end{aligned} \tag{3.226}$$

$$\begin{aligned}
& \oint \frac{dw}{2\pi i} j_1(w) \xi(z) c(z) V(z) e^{-\phi(z)} = \\
& = \frac{1}{2} \oint \frac{dw}{2\pi i} \left(e^{\phi(w)-\chi(w)} e^{-\phi(z)+\chi(z)} \right) c(z) \left(T_F(w) V(z) \right) = \\
& = \frac{1}{2} \oint \frac{dw}{2\pi i} \left(1 + \dots \right) c(z) \left(\frac{1}{w-z} V_B(z) + \dots \right) = \frac{1}{2} c(z) V_B(z),
\end{aligned} \tag{3.227}$$

where we used the SCFT OPE

$$T_F(z) V(0) \sim \frac{1}{z} V_B(0) \tag{3.228}$$

for the fermionic $h = \frac{1}{2}$ superconformal field

$$V(z, \theta) = V(z) + \theta V_B(z). \tag{3.229}$$

Thus the picture changing acts on NS vertices as

$$c V_{(-1)}(z) \equiv c e^{-\phi} V(z) \longrightarrow c V_{(0)}(z) \equiv c V_B(z) - \frac{1}{2} \gamma V(z). \tag{3.230}$$

It is customary to write the picture-zero vertex without the last term, since it will not contribute in essentially all amplitudes by γ -number conservation. Then we get for the zero-picture vertex just

$$c V_{(0)}(z) = c V_B(z) \equiv c \int d\theta V(z, \theta), \tag{3.231}$$

vindicating the identification $e^{-\phi} \approx \delta(\gamma) \leftrightarrow \int d\theta$ discussed around Eq. (3.109).

Triviality of BRST Cohomology in $\mathcal{A}_{\text{large}}$ -Modules

We check that, as claimed, the Q -cohomology is trivial in the large algebra. One has

$$1 = XY = 2\{Q, \xi\}Y = 2Q\xi Y + 2\xi QY = 2Q\xi Y + 2\xi YQ = \{Q, 2\xi Y\}, \tag{3.232}$$

so that in the large algebra the identity is a BRST-homotopy. This entails that the cohomology is trivial: indeed for all Q -closed state $|\psi\rangle$, we have

$$|\psi\rangle = 1 \cdot |\psi\rangle = 2\{Q, \xi Y\}|\psi\rangle = Q(2\xi Y|\psi\rangle), \tag{3.233}$$

so that $|\psi\rangle$ is automatically Q -exact. We stress that ξY is not a legitimate operator in the small algebra, so the state $|\psi\rangle$ is not Q -exact in the “Hilbert space” $\mathbf{H}_{\text{small}}$ isomorphic to the small operator algebra $\mathcal{A}_{\text{small}}$ via the CFT state/operator isomorphism. We stated after Eq. (3.187) that in the large algebra, while $Q|0\rangle = 0$, the Hermitian conjugate of $|0\rangle$ is not Q -closed. We now can make the statement more precise:

$$Q(|0\rangle)_{\text{large}}^\dagger \equiv Q|\xi\rangle = Q\xi|0\rangle = \{Q, \xi\}|0\rangle = \frac{1}{2}|X\rangle \neq 0. \quad (3.234)$$

One could have wondered why we took $\mathcal{A}_{\text{small}}$, and not $\mathcal{A}_{\text{large}}$, as the physical operator algebra. The short answer is that $\mathbf{H}_{\text{small}}$ is the largest space in which Q acts as a unitary operator, as required for a consistent BRST quantization.

Note 3.7 Things look simpler from the path integral viewpoint. The large-system path integral vanishes because of the ξ zero-mode unless we insert a ξ to soak it

$$\int d[\text{large system fields}] e^{-S} \xi(z) O_1(z_1) O_2(z_2) \dots O_s(z_s). \quad (3.235)$$

$\xi(z)$ is not Q -closed, and the amplitude is not defined in large Q -cohomology.

3.7.2 BRST Cohomology in Operator Space: Vertices

Next we consider Q -cohomology valued in operator spaces. By state-operator correspondence, this cohomology is canonically isomorphic to the one for states. The homotopy arguments continue to work: for instance, since $[Q, L_0] = 0$, we may compute the Q -cohomology in the space of operators having definite conformal weight $[L_0, O] = h O$. If O is Q -closed,

$$h O = [[Q, b_0], O] = [[b_0, O], Q] \quad (3.236)$$

and hence all Q -cohomology classes have a representative with $h = 0$. The same argument with the replacement $L_0 \rightsquigarrow E + N_g$, shows that all picture (-1) NS Q -cohomology classes have a representative of the form

$$c V e^{-\phi}, \quad \left(V \text{ an } h = \frac{1}{2} \text{ matter superconformal primary} \right), \quad (3.237)$$

while the picture $(-\frac{1}{2})$ R-sector Q -cohomology classes are represented by operators of the form

$$c \left((\text{polynomial in } \partial^k X^\mu, \partial^\ell \psi^\nu) S e^{ik \cdot X} \right) e^{-\phi/2}, \quad (3.238)$$

where S is a $SO(9, 1)$ spin field (S_α or $S_{\dot{\alpha}}$), and the matter operator inside the large parenthesis has $h = 5/8$ and (anti)commutes with G_r , $r \geq 0$. Of course, these results

can also be obtained as an elementary application of the state-operator correspondence.

Let us check these facts and their consequences in detail. In Eq. (3.160), we decomposed the BRST charge Q in three pieces Q_0 , Q_1 , and Q_2 according to their β , γ charge. Let $j_0(z)$ be the supercurrent corresponding to charge Q_0 which has zero β , γ ghost number; see Eq. (3.161). Since Q_0 has the same form as the world-sheet reparametrization BRST charge for the bosonic string, we can borrow the following formulae from bosonic string theory; they are valid for all operators $O(z)$ containing matter fields and β , γ ghosts (equivalently, containing matter fields and ϕ , ξ , and η):

$$\oint \frac{dw}{2\pi i} j_0(w) O(z) = h(\partial c)O(z) + c\partial O(z) = (h-1)(\partial c)O(z) + \partial(cO(z)) \quad (3.239)$$

$$\Rightarrow \oint \frac{dw}{2\pi i} j_0(w) c(z) O(z) = (h-1) : \partial c c : O(z), \quad (3.240)$$

where h is the conformal weight of $O(z)$. $j_1(z)$ is proportional to

$$\gamma(z) T_F^{\text{matt}}(z) = e^{\phi(z)} \eta(z) T_F^{\text{matt}}(z), \quad (3.241)$$

and $j_2(z)$ is proportional to $\gamma(z)^2 b(z)$. Their OPEs with a NS operator in picture $-q$, i.e. of the form $cV e^{-q\phi}$, with V a matter operator, are

$$\begin{aligned} e^{\phi(z)} \eta(z) T_F^{\text{matt}}(z) c(w) V(w) e^{-q\phi(w)} &= \\ &= c(w) \left((z-w)^q + \dots \right) \left(T_F^{\text{matt}}(z) V(w) \right) e^{(1-q)\phi}. \end{aligned} \quad (3.242)$$

★ Changing Picture in Integrated Vertices

The picture changing above is for *local* BRST-invariant operators $cV(z)$. In string theory we are also interested in BRST-invariants of the form $\int dz V(z)$. (Integrated) physical vertex operators are of this form, and only the integral is required to be BRST-invariant, that is, $[Q, V(z)]$ needs not to vanish; it may be just a total derivative. This means that $[Q, \xi V]$ contains a total derivative of the form $\partial(\xi cV)$ which is not part of the “small” operator algebra, and hence its presence is not legitimate. To fix this problem, one defines the picture changing for integrated vertices as

$$V_{(q)} \rightarrow V_{(q+1)} = 2[Q, \xi V_{(q)}] + 2\partial(\xi cV_{(q)}). \quad (3.243)$$

Picture Changing for the R Vertex

The R vertex corresponding to the massless fermions (the “Ramond vacua”) in the standard picture $-\frac{1}{2}$ reads

$$V_{(-1/2)}(z) = c u^\alpha S_\alpha(z) e^{ik \cdot X} e^{-\phi(z)/2}, \quad k^2 = \not{k}u = 0, \quad (3.244)$$

where the on-shell conditions

$$k^2 = 0 \quad \text{and} \quad \not{k}u = 0 \quad (3.245)$$

follow from $[Q_0, V_{(-1/2)}] = 0$, and $[Q_1, V_{-1/2}] = 0$, respectively.

Let us find the picture $(+\frac{1}{2})$ vertex. In the picture changing we may ignore Q_0 , which gives just a total derivative (cf. Eq. (3.239)), while Q_2 produces a term proportional to b which is usually ignored since it does not contribute to amplitudes. The essential term is the one which arises from Q_1 , i.e. the term $e^\phi T_F$ in the X operation; cf. Eq. (3.204). For the example where the “matter” is the free theory (2.397), we have

$$\begin{aligned} V_{(+1/2)}(z) &= X V_{(-1/2)}(z) = c \lim_{w \rightarrow z} e^{\phi(w)} \psi^\mu(w) \partial X_\mu(w) u^\alpha S_\alpha(z) e^{ik \cdot X(z)} e^{-\phi(z)/2} + \dots = \\ &= c \lim_{w \rightarrow z} \left(e^{\phi(w)} e^{-\phi(z)/2} \right) \left(u^\alpha \psi^\mu(w) S_\alpha(z) \right) \left(\partial X_\mu(w) e^{ik \cdot X(z)} \right) + \dots = \\ &= c \lim_{w \rightarrow z} \left((w-z)^{1/2} e^{\phi(w)-\phi(z)/2} \right) \left(u^\alpha (w-z)^{-1/2} (\gamma^\mu)_\alpha^{\dot{\beta}} S_{\dot{\beta}}(z) + O((w-z)^{1/2}) \right) \times \\ &\quad \times \left(\frac{ik_\mu}{w-z} e^{ik \cdot X} + \partial X_\mu e^{ik \cdot X} + O(w-z) \right) + \dots \end{aligned}$$

Note that the pole term cancels by the on-shell (or BRST) condition $k u = 0$. However, the $O((w-z)^{1/2})$ sub-leading term in the OPE $\psi^\mu(w) S_\alpha(z)$ gives a finite contribution. We look to this sub-leading operator which, by $SO(10)$ symmetry and dimension considerations, should have the structure

$$k_\mu (M^{\mu\rho\sigma})_\alpha^{\dot{\beta}} \psi_\rho \psi_\sigma S_{\dot{\beta}}, \quad (3.246)$$

for some Lorentz intertwiner $(M^{\mu\rho\sigma})_\alpha^{\dot{\beta}}$ antisymmetric in $\rho\sigma$. So

$$(M^{\mu\rho\sigma})_\alpha^{\dot{\beta}} = c_1 \eta^{\mu[\rho} (\gamma^{\sigma]}_\alpha^{\dot{\beta}} + c_2 (\gamma^\mu \gamma^{[\rho} \gamma^{\sigma]})_\alpha^{\dot{\beta}} \quad (3.247)$$

for certain coefficients c_1, c_2 . These coefficients may be computed in various ways, e.g. using the bosonized form of the $SO(10)$ current algebra. The second term produces a contribution proportional to $k u = 0$, so only the first term matters. We are left with

$$V_{(1/2)}(z) = c e^{\phi/2} u^\alpha (\gamma^\mu)_\alpha^{\dot{\beta}} \left(\partial X_\mu + i c_1 k \cdot \psi \psi_\mu \right) S_{\dot{\beta}} e^{ik \cdot X} + \dots, \quad (3.248)$$

where $\dots$ is the omitted term proportional to b . By bosonization one finds $c_1 = \frac{1}{4}$.

The $-\frac{3}{2}$ Picture Vertex

Recall from chapter 2 that the dimension of the primary operator $e^{q\phi(z)}$

$$h_q = -\frac{1}{2}q(q+2) \quad (3.249)$$

is invariant under $q \leftrightarrow -(q+2)$. Hence $e^{-\phi/2}$ and $e^{-3\phi/2}$ have the same dimension $\frac{3}{8}$ and $q = -1/2$ and $-3/2$ are “equally standard” choices of the picture.³¹ Thus the two operators

³¹ Note that the NS standard picture $q = -1$ is invariant under the $q \leftrightarrow -(q+2)$ reflection.

$$S_\alpha e^{-\phi/2} \quad \text{and} \quad S_{\dot{\alpha}} e^{-3\phi/2} \quad (3.250)$$

have the same dimension, $h = 1$, and are both GSO allowed in $R-$ (cf. Sect. 3.1). The first operator, multiplied by the obvious Dirac wave-function factor $u^\alpha e^{ik \cdot X}$ is the *standard* picture $(-\frac{1}{2})$ massless fermionic vertex (which is the picture which makes the comparison with light-cone/OCQ quantizations more direct)

$$V_{(-1/2)}(u, k) = c u^\alpha S_\alpha e^{-\phi/2} e^{ik \cdot X(z)}. \quad (3.251)$$

One would like to identify

$$V_{(-3/2)}(v, k) \equiv c v^{\dot{\alpha}} S_{\dot{\alpha}} e^{-3\phi/2} e^{ik \cdot X(z)} \quad k^2 = 0 \quad (3.252)$$

with the picture $(-\frac{3}{2})$ vertex. From the above dimensional considerations we have $[Q_0, V_{(-3/2)}] = 0$, and it is also obvious that $[Q_2, V_{(-3/2)}] = 0$. On the other hand,

$$\begin{aligned} e^\phi(z) \eta(z) \psi^\mu(z) \partial X_\mu(z) v^{\dot{\alpha}} S_{\dot{\alpha}}(w) e^{-3\phi(w)/2} e^{ik \cdot X(w)} \sim \\ \sim \left((z-w)^{3/2} e^{-\phi(w)/2} + \dots \right) \left(\eta(w) + \dots \right) \times \\ \times \left(\frac{ik_\mu}{z-w} e^{ik \cdot X(w)} + \dots \right) \left(\frac{v^{\dot{\alpha}} (\gamma^\mu)_{\dot{\alpha}}^\alpha S_\alpha(w)}{(z-w)^{1/2}} + \dots \right) \end{aligned} \quad (3.253)$$

is non-singular, and hence $[Q_1, V_{(-3/2)}] = 0$ for *all values* of the coefficient $v^{\dot{\alpha}}$ (at $k^2 = 0$). Multiplying both sides of Eq. (3.253) by $\xi(w)$ replaces the second parenthesis in the RHS by $(1/(z-w) + \eta\xi)$

$$2\{Q, \xi V_{(-3/2)}(z)\} = i v^{\dot{\alpha}} (k)_{\dot{\alpha}}^\alpha S_\alpha e^{-\phi/2} e^{ik \cdot X(z)} \quad (3.254)$$

so the picture changing of $V_{(-3/2)}(z)$ is the picture $(-\frac{1}{2})$ vertex $V_{(-1/2)}(z)$ with $u = ikv$

$$X V_{(-3/2)}(v, k) = V_{(-1/2)}(ikv, k). \quad (3.255)$$

Since $k^2 = 0$, we have

$$ku = i(k)^2 v = ik^2 v \equiv 0 \quad (3.256)$$

identically for all v 's. Thus the $(-1/2)$ vertex is automatically on-shell (for $k^2 = 0$) consistently with the fact that $V_{(-3/2)}(v, k)$ was BRST-invariant for all v .

3.7.3 RR Vertices and a Perturbative Theorem

In the closed superstring RR states are bosonic. The vertices for these bosonic states are given by the product of one R vertex from the left- and one from the right-movers.

The structure of the resulting bosonic vertex may be confusing at first, so we proceed at a slow pace. Consider, say, the sector $(R+, R+)$ where we perform the *same* GSO projection on both sides: the massless particles in this sector are listed in the second row of Table 3.1. That table was obtained in the framework of light-cone quantization: the $(R+, R+)$ massless physical states correspond to the product of a left-moving and a right-moving Ramond ground state of the *transverse matter system* with chirality $+$. Therefore they make the representation $S_+ \otimes S_+$ of $\text{Spin}(8)$, where S_+ is the chirality $+$ spin representation. $S_+ \otimes S_+$ decomposes as

$$S_+ \otimes S_+ = \mathbf{1} \oplus \Lambda_2 \oplus \Lambda_4^+; \quad (3.257)$$

see **BOX 3.1**. Therefore the massless $(R+, R+)$ sector consists of a scalar, a 2-form field, and a self-dual 4-form field.

Now let us go to the covariant BRST quantization. In (say) the $(-\frac{1}{2}, -\frac{1}{2})$ -picture the vertex of a massless $(R-, R-)$ state should be proportional to

$$S_\alpha(z) \tilde{S}_\beta(\bar{z}) e^{-(\phi+\tilde{\phi})/2} e^{ik \cdot X(z, \bar{z})} \quad k^2 = 0 \quad (3.258)$$

so now it is in the $S_+ \otimes S_+$ representation of $SO(10)$. Since $SO(10)$ is of the form $SO(4n+2)$, and not of the form $SO(4n)$, from the same **BOX 3.1** we see that its decomposition into irreducible representations contains forms of the opposite parity, i.e. *odd* forms instead of even

$$S_+ \otimes S_+ = \Lambda_1 \oplus \Lambda_3 \oplus \Lambda_5^+ \equiv \mathbf{10} \oplus \mathbf{120} \oplus \mathbf{126}, \quad (3.259)$$

which is surprising at first since we know from the light-cone analysis that the physical $(R-, R-)$ spectrum contains a scalar, a 2-form, and a (self-dual) 4-form, whereas the covariant vertices transform as a vector, a 3-form, and a (self-dual) 5-form. *How do we reconcile this mismatch of Lorentz representations?*

The crucial observation is that only one-quarter (i.e. 64 out of 256) operators in the representation (3.259) are on-shell and hence BRST-invariant. To get a physical vertex, we need to contract the free spinor indices α, β with a matrix $M^{\alpha\beta}$ which projects the operator into its BRST-invariant part (with respect to both left and right BRST charges $Q, \tilde{Q}$). The BRST invariance conditions with respect to Q and $\tilde{Q}$ are, respectively, the left- and right- Dirac equations (cf. Eq. (3.244))

$$M\mathcal{H} = \mathcal{H}M = 0 \quad \text{for } k^2 = 0. \quad (3.260)$$

The solution to the BRST conditions (3.260) is given by Clifford-algebra matrices of the form

$$M = k_\mu \epsilon_{\mu_1 \dots \mu_s} \gamma^{\mu \mu_1 \dots \mu_s} \quad \text{with } \epsilon_{\mu_1 \dots \mu_s} \begin{cases} \text{totally antisymmetric and} \\ \underline{\text{transverse}} k^{\mu_1} \epsilon_{\mu_1 \dots \mu_s} = 0. \end{cases} \quad (3.261)$$

We interpret the constant tensors $\epsilon_{\mu_1 \dots \mu_s}$ as the polarizations of the RR particles (up to overall normalization). Note that a polarization of the form

$$\epsilon_{\mu_1 \dots \mu_s} = k_{[\mu_1} \lambda_{\mu_2 \dots \mu_s]} \quad \text{with } \lambda_{\mu_1 \dots \mu_{s-1}} \text{ totally antisymmetric} \quad (3.262)$$

yields a vanishing M , and hence the transformation

$$\epsilon_{\mu_1 \dots \mu_s} \rightarrow \epsilon_{\mu_1 \dots \mu_s} + k_{[\mu_1} \lambda_{\mu_2 \dots \mu_s]} \quad \lambda_{\mu_2 \dots \mu_s} \text{ arbitrary} \quad (3.263)$$

does not change the physical S -matrix. We conclude that the RR massless states are spacetime s -form gauge fields A , whose gauge transformations have the form

$$A \rightarrow A + d\lambda \quad (3.264)$$

with λ an arbitrary $(s-1)$ -form gauge parameter. The gauge-invariant field strength $(s+1)$ -forms are then (at the linearized level)

$$F = dA. \quad (3.265)$$

Because of the independent chirality projections (GSO projections) on the left- and the right side, the spacetime $(s+1)$ -form

$$\tilde{S} \gamma_{\mu\mu_1 \dots \mu_s} S \quad (3.266)$$

is non-zero only for s even (resp. odd) in $(R\pm, R\pm)$ (resp. $(R\pm, R\mp)$) and is (anti)self-dual since

$$\gamma_{\mu_1 \dots \mu_{s+1}} \gamma_{11} = \frac{1}{(10-s-1)!} \epsilon_{\mu_1 \dots \mu_{s+1} \mu_{s+2} \dots \mu_{10}} \gamma^{\mu_{s+2} \dots \mu_{10}}, \quad (3.267)$$

while the GSO projection implies

$$\gamma_{11} S = \mp S \quad \text{in the } R\pm \text{ sector.} \quad (3.268)$$

So, say for $(R-, R-)$, the independent polarizations are

$$\epsilon, \quad \epsilon_{\mu_1 \mu_2}, \quad \epsilon_{\mu_1 \mu_2 \mu_3 \mu_4}, \quad (3.269)$$

subjected to the transversality and anti-self-duality conditions

$$k^{\mu_1} \epsilon_{\mu_1 \dots \mu_s} = 0 \quad \text{for } s > 0 \quad (3.270)$$

$$k_{[\mu_1} \epsilon_{\mu_2 \mu_3 \mu_4 \mu_5]} = \frac{1}{5!} \epsilon_{\mu_1 \dots \mu_{10}} k^{\mu_6} \epsilon^{\mu_7 \mu_8 \mu_9 \mu_{10}} \quad \text{for } s = 4, \quad (3.271)$$

and identified modulo the gauge equivalence (3.263). It is easy to see that we get the correct physical degrees of freedom as found, e.g. in the light-cone quantization.

In conclusion, we see that the RR vertex may be written as

$$c_s \, c\tilde{C} \, F(X)_{\mu_1 \dots \mu_{s+1}} \, \tilde{S} \gamma^{\mu_1 \dots \mu_{s+1}} S \, e^{-(\phi + \tilde{\phi})/2}, \quad (3.272)$$

where $F(X)_{\mu_1 \dots \mu_{s+1}}$ has the physical interpretation of being the *linearized-level, on-shell field strength* $(s+1)$ form $F = dA$. In (3.272), c_s is a numerical normalization coefficient which is easily determined by computing the two-point functions; if the spin fields S are normalized in the standard way, one gets $c_s = 1/(4\sqrt{s!})$.

From Eq. (3.272) we learn several lessons.

First of all, the spacetime field strength 5-form of the 4-form gauge field in the $(R+, R+)$ sector satisfies a *self-duality constraint* that makes it very subtle (e.g. it cannot be described by a conventional Lagrangian).

On the other hand, all RR vertices depend directly on the field strength rather than the gauge field. This means that no *perturbative* state of the superstring can carry a non-zero charge under a RR gauge symmetry. Indeed the charge $q_{\mu_1 \dots \mu_{s-1}}$ of a state $|a, p\rangle$ (of momentum p) under a s -form gauge field $A_{\mu_1 \dots \mu_s}$ of vertex $V(k)_{\mu_1 \dots \mu_s}$ is

$$q_{\mu_1 \dots \mu_{s-1}} = \lim_{k \rightarrow 0} \langle a, p + k | V(k)_{\mu_1 \dots \mu_{s-1} \mu_0} | a, p \rangle, \quad (3.273)$$

which vanishes in the RR case since the vertex is proportional to k . Equivalently, from the form of the RR vertex we conclude that the low-energy effective Lagrangian of the superstring contains the RR gauge forms A only through their field strengths F , and no field-current coupling $A \cdot J$ is present. In particular, the 0-form in the $(R+, R+)$ sector is a *Peccei-Quinn* scalar, that is, a scalar a endowed with a symmetry

$$a \rightarrow a + \text{const.} \quad (3.274)$$

also known as an *axion*.³² We stress that this conclusion—being obtained in the framework of string *perturbation* theory—holds only if $|a, p\rangle$ is a *perturbative state* of the superstring, namely a state whose mass remains bounded as the string coupling constant vanishes, $g_s \equiv e^\Phi \rightarrow 0$. We shall see in Chap. 12 that there are non-perturbative objects in superstring theory which *do carry non-zero RR charges*.

The $(-1/2, -3/2)$ RR Vertex: The Dirac-Kähler Equation

For certain applications one needs the RR vertex for the RR gauge potentials A themselves, rather than for their field strengths $F = dA$. This can be achieved using the fermion vertex in picture $-\frac{3}{2}$. We consider the RR vertex in the *asymmetric* picture $(-\frac{1}{2}, -\frac{3}{2})$, i.e.

$$\Lambda^{\alpha\tilde{\beta}} (S_\alpha e^{-\phi/2}) (\tilde{S}_{\tilde{\beta}} e^{-3\tilde{\phi}/2}) e^{ik \cdot X(z, \bar{z})} \quad k^2 = 0. \quad (3.275)$$

³² As we shall see, a is a spacetime *pseudo*-scalar.

Since on the left we have the standard $(-\frac{1}{2})$ picture, as before, invariance under the *left* BRST symmetry Q requires

$$\not{K} \Lambda = 0. \quad (3.276)$$

But now on the right we have picture $(-\frac{3}{2})$ and it follows from Eqs. (3.252)–(3.256) that *no* further condition on Λ is imposed by the right BRST symmetry $\tilde{Q}$. Hence we get a condition on Λ which is much weaker than the one we got for M in the symmetric $(-\frac{1}{2}, -\frac{1}{2})$ picture, Eq. (3.260). In **BOX 3.3** it is shown that this condition is exactly the (*Dirac*)–Kähler equation

$$(d - \delta)\Lambda = 0. \quad (3.277)$$

By performing a picture changing on the right, we get back the vertex in the symmetric picture $(-\frac{1}{2}, -\frac{1}{2})$ with

$$M = (d + \delta)\Lambda = d(2\Lambda). \quad (3.278)$$

Since M is the (linearized) field strength (up to the overall normalization, which we have *not* kept track of), we see that the Kähler form field Λ is the gauge potential (up to normalization). Equation (3.277) is then a “generalized” Lorentz gauge for this gauge potential.

In a sense the $(-1/2, -3/2)$ vertex is more fundamental than the standard one. Recall that the BRST cohomology is isomorphic to the light-cone Hilbert space only at zero-momentum. There are BRST-invariant configurations at $p_\mu = 0$ which are not the $p_\mu \rightarrow 0$ limit of the ones at $p_\mu \neq 0$. These quasi-topological d.o.f. are most easily analyzed using the $(-1/2, -3/2)$.

Disk Tadpoles for RR Vertices

We explain why the asymmetric $(-1/2, -3/2)$ is more “fundamental”.³³ Consider the would-be tree-level emission of a RR state in the open oriented superstring.³⁴ The world-sheet is a disk, and we have to insert a RR vertex which has left–right pictures $(k + \frac{1}{2}, k' + \frac{1}{2})$, for some integers k and k' . However, the total left+right picture of the insertions inside a disk (recall that the boundary condition identifies left- and right-moving d.o.f.) is equal to minus the number of γ zero-modes, i.e. -2

$$\left(k + \frac{1}{2}\right) + \left(k' + \frac{1}{2}\right) = -2 \quad \Rightarrow \quad k + k' = -3, \quad (3.279)$$

³³ More conceptually, an inversion of the orientation of the world-sheet, which interchanges left- and right-movers, is naturally associated with an interchange of the ghost sea level with its dual one.

³⁴ Physically this theory does not exist (see Chap. 5). Our discussion is purely mathematical.

BOX 3.3 - The Kähler-Dirac equation & BRST-invariant RR vertices

We look for the solutions to

$$\mathcal{H}\Lambda = 0, \quad (\star)$$

where Λ is an element of the Clifford algebra in d dimensions $\mathbb{C}\ell(d)$. The Clifford algebra is $\mathbb{Z}_2$ graded, $\mathbb{C}\ell(d) = \mathbb{C}\ell(d)_0 \oplus \mathbb{C}\ell(d)_1$, and we may assume with no loss that Λ has a definite parity, *even* or *odd*. Expanding in the usual basis of $\mathbb{C}\ell(10)_{0,1}$, we see that a general Λ has the form

$$\Lambda = \sum_{\mu_1 \dots \mu_k} \Lambda_{\mu_1 \dots \mu_k}^{(k)} \gamma^{\mu_1 \dots \mu_k} \quad (\star\star)$$

and

$$\begin{aligned} \mathcal{H}\Lambda &= k_\mu \sum_k \Lambda_{\mu_1 \dots \mu_k}^{(k)} (\eta^{\mu[\mu_1} \gamma^{\mu_2 \dots \mu_k]} + \gamma^{\mu\mu_1 \dots \mu_k}) = \\ &= \sum_k \left(k^\mu \Lambda_{\mu\mu_1 \dots \mu_{k-1}}^{(k)} \gamma^{\mu_1 \dots \mu_{k-1}} + k_{\mu_1} \Lambda_{\mu_2 \dots \mu_{k+1}}^{(k)} \gamma^{\mu_1 \dots \mu_{k+1}} \right). \end{aligned}$$

Hence,

$$-\delta\Lambda^{(k+2)} + d\Lambda^{(k)} = 0 \quad (\star\star\star)$$

where $\delta = - * d * \equiv d^\dagger$. Thus Λ , identified with the differential form (of mixed degree *but* definite degree parity)

$$\Lambda \longleftrightarrow \sum_{k \text{ even(odd)}} \Lambda_{\mu_1 \dots \mu_k}^{(k)} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_k},$$

satisfies Eq. ($\star\star\star$) known as the *Kähler equation* (a relative to Dirac equation)

$$(d - \delta)\Lambda = 0 \Rightarrow 0 = -(d - \delta)^2 \Lambda = (\delta d + d\delta)\Lambda \equiv \Delta\Lambda,$$

where Δ is the *Hodge Laplacian* (recall that $\delta^2 = d^2 = 0$). The Clifford $\mathbb{Z}_2$ degree plays for the Kähler equation the same role as chirality for the usual Dirac equation. From Eq. ($\star$), we see that if Λ is a solution to the Kähler equation so is $\gamma_{11}\Lambda$. So it makes sense to impose the GSO projections $\gamma_{11}\Lambda = \pm\Lambda$, which relate the form-coefficients in Eq. ($\star\star$) in the form $\Lambda^{(10-k)} = \pm\Lambda^{(k)}$, so that only the $k \leq 4$ ones are really independent.

which has no symmetric solution $k = k' \in \mathbb{Z}$. The most economic (and canonical) solution is $k = -1$, $k' = -2$ which leads to RR vertices of picture $(-1/2, -3/2)$.

The example of the disk tadpole illustrates why BRST-invariant configurations frozen at $p_\mu = 0$ (with no counterpart in the light-cone) may play an important role in physical amplitudes. See Chap. 5 for more.

Quasi-Topological Degrees of Freedom

States which exist only at $p_\mu = 0$ are quasi-topological, i.e. they do not propagate local d.o.f. In d dimensions there are two kinds of quasi-topological gauge fields, the d -form ones $A^{(d)}$ with the gauge-invariant action

$$\mu \int_{M_d} A^{(d)} \quad (3.280)$$

for some coupling constant μ . The other one is a gauge $(d-1)$ -form with field strength $F^{(d)} = dA^{(d)}$ and Maxwell action

$$\frac{1}{2d! g^2} \int_{M_d} F^{(d)} \wedge *F^{(d)}. \quad (3.281)$$

Exercise 3.3 Show that the Lagrangians (3.280), (3.281) describe non-propagating zero-momentum d.o.f.

The BRST analysis of the $(-1/2, -3/2)$ RR vertex shows that the non-propagating physical mode $A^{(10)}$ (resp. $A^{(9)}$) exists in the $(R+, R+)$ (resp. $(R+, R-)$) sector.

3.8 Spacetime Supersymmetry

We recall how the ordinary spacetime Poincaré group is realized in the superstring. The string is a theory of quantum gravity, so Poincaré invariance is part of the gravitation Ward identities. More directly, the Poincaré generators are the conserved charges associated (via the Noether theorem) with the Poincaré symmetries of the world-sheet action (2.397). Let us focus on the spacetime translation operator P^μ . To this generator there corresponds on the world-sheet a conserved current which is physical, i.e. BRST-invariant and not BRST trivial (otherwise it would vanish on all physical states). The left-moving part of the translation current is then a physical vertex at zero-momentum, i.e. the $k_\mu \rightarrow 0$ limit of the massless NS vertex. In the zero picture we get ∂X^μ , which coincides with the usual Noether current of the shift symmetry $X^\mu \rightarrow X^\mu + a^\mu$. In picture -1 this becomes the current $\psi^\mu e^{-\phi}$

$$P_{(-1)}^\mu = \oint \frac{dz}{2\pi i} \psi^\mu(z) e^{-i\phi(z)} + \text{right-movers}. \quad (3.282)$$

From Eq. (3.243)

$$P_{(0)}^\mu \equiv \oint \frac{dz}{2\pi i} \partial X^\mu(z) = 2 \left[Q, \oint \frac{dz}{2\pi i} \xi(z) \psi^\mu(z) e^{-i\phi(z)} \right]. \quad (3.283)$$

It follows from the rearrangement lemma (3.221) that the insertion of the picture (-1) version of P_μ in a BRST-invariant amplitude is equivalent to the introduction of the Noether charge $\oint dz \partial X_\mu / (2\pi i)$ (while preserving the condition (3.220)).

The situation with the spacetime SUSY supercharge is analogous. The only difference is that the supercharge must belong to the R-sector since it transforms bosons into fermions. Then the world-sheet current associated with a spacetime supercharge should be the massless R vertex at $k = 0$. In the usual $(-\frac{1}{2})$ picture

$$Q_\alpha = \oint \frac{dz}{2\pi i} S_\alpha(z) e^{-\phi(z)/2}, \quad (3.284)$$

while in the $(+\frac{1}{2})$ picture it is

$$Q_\alpha = \oint \frac{dz}{2\pi i} (\partial X_\mu)(\gamma^\mu)_\alpha^{\dot{\beta}} S_{\dot{\beta}}(z) e^{\phi(z)/2}. \quad (3.285)$$

Let us check the superPoincaré algebra

$$[P_\mu, P_\nu] = 0, \quad [P_\mu, Q_\alpha] = 0, \quad \{Q_\alpha, Q_\beta\} = (\gamma^\mu)_{\alpha\beta} P_\mu. \quad (3.286)$$

The first equation follows from the OPEs

$$\text{pictures } 0,0 \quad \partial X_\mu(z) \partial X_\nu(w) \sim \frac{\eta^{\mu\nu}}{(z-w)^2} \quad (3.287)$$

$$\text{pictures } 0,1 \quad \partial X_\mu(z) \psi_\nu(w) e^{-\phi(w)} = \text{regular}. \quad (3.288)$$

Note that the Poincaré “current algebra” takes different forms in different pictures, even if the physics is picture-independent. The second Eq. (3.286) follows from

$$\partial X_\mu(z) S_\alpha(w) e^{-\phi(w)/2} = \text{regular as } z \rightarrow w. \quad (3.289)$$

The third Eq. (3.286) follows most easily from supercharges in pictures $(+\frac{1}{2})$ and $(-\frac{1}{2})$

$$\begin{aligned} \partial X_\mu(z) (\gamma^\mu)_\alpha^{\dot{\beta}} S_{\dot{\beta}}(z) e^{\phi(z)/2} S_\gamma(w) e^{-\phi(w)/2} &= \\ &= \partial X_\mu(z) (\gamma^\mu)_\alpha^{\dot{\beta}} \left(e^{\phi(z)/2} e^{-\phi(w)/2} \right) \left(S_{\dot{\beta}}(z) S_\gamma(w) \right) = \\ &= \partial X_\mu(z) (\gamma^\mu)_\alpha^{\dot{\beta}} \left((z-w)^{1/4} + \dots \right) \left((z-w)^{-5/4} C_{\dot{\beta}\gamma} + \dots \right) \sim \\ &\sim \frac{1}{z-w} \partial X_\mu (\gamma^\mu)_{\alpha\gamma} \end{aligned} \quad (3.290)$$

where we used (2.493).

The spacetime supercharges map bosonic physical states (having integral space-time) into fermionic physical states (spacetime spinors), and vice versa. By the CFT state-operator correspondence this also holds for the corresponding vertices

$$\oint_C \frac{dw}{2\pi i} Q_\alpha(w) V(z)_{\text{Bose}} = V(z)_{\text{Fermi}}, \quad (3.291)$$

where $Q_\alpha(w)$ stands for the zero-momentum massless R vertex in some chosen picture. In Eq. (3.291), the picture of the resulting fermionic vertex is the sum of the pictures for the bosonic vertex and the supercharge Q_α .

Example: SUSY transformation of a massless boson

We check that Q_α maps a massless Bose vertex into a Fermi one. In pictures (-1) , $(-\frac{1}{2})$ one has

$$\begin{aligned} \left[Q_\alpha, \psi^\mu(w) e^{ik \cdot X(w)} e^{-\phi(w)} \right] &= \oint \frac{dz}{2\pi i} S_\alpha(z) e^{-\phi(z)/2} \psi^\mu(w) e^{ik \cdot X(w)} e^{-\phi(w)} = \\ &= \oint \frac{dz}{2\pi i} \left((w-z)^{-1/2} e^{-\phi(z)/2 - \phi(w)} \right) \left((w-z)^{-1/2} (\gamma^\mu)_\alpha{}^\beta S_\beta \right) e^{ik \cdot X} = e^{-3\phi/2} (\gamma^\mu)_\alpha{}^\beta S_\beta e^{ik \cdot X}, \end{aligned}$$

which is the picture $(-\frac{3}{2})$ vertex. In the (-1) , $(+\frac{1}{2})$ picture one gets

$$\begin{aligned} \left[Q_\alpha, \psi^\mu(w) e^{ik \cdot X(w)} e^{-\phi(w)} \right] &= \oint \frac{dz}{2\pi i} \partial X_\nu(z) (\gamma^\nu)_\alpha{}^\beta S_\beta(z) e^{+\phi(z)/2} \psi^\mu(w) e^{ik \cdot X(w)} e^{-\phi(w)} = \\ &= \oint \frac{dz}{2\pi i} (w-z)^{1/2} e^{-\phi/2} \frac{ik_\nu}{z-w} (w-z)^{-1/2} (\gamma^\nu \gamma^\mu)_\alpha{}^\beta S_\beta e^{ik \cdot X} = e^{-\phi/2} (\gamma^\mu \gamma^\nu)_\alpha{}^\beta S_\beta e^{ik \cdot X}, \end{aligned}$$

which is the Fermi vertex in picture $(-\frac{1}{2})$.

3.8.1 Supersymmetry Ward Identities: Absence of Tadpoles

The superstring contains propagating massless spin $3/2$ particles. Non-free massless spin- $\frac{3}{2}$ particles have a consistent propagation if and only if they are the gravitini of a supergravity [29], i.e. if they are the gauge particles of a *local* supersymmetry invariance. Thus the presence of gravitino requires a complete set of SUSY Ward identities. Let us show how the very existence of a massless Fermi vertex implies these identities.

We saw above that the SUSY transformation of a vertex V takes the form

$$\delta_\epsilon V(z) = \oint \frac{dw}{2\pi i} \epsilon^\alpha Q_\alpha(w) V(z), \quad (3.292)$$

where $Q_\alpha(w)$ is the massless spinor vertex at $k=0$ in a suitable picture and ϵ^α a Grassmann parameter. *Let us assume that*

$$s(w) := \epsilon^\alpha Q_\alpha(w) \frac{dw}{2\pi i} \quad (3.293)$$

is a well-defined global 1-form on the (oriented) world-sheet Σ . Being holomorphic, s is in particular closed $ds=0$. Let D_j be a small disk centered at the point z_j and $C_j = \partial D_j$ be the small circle at its boundary. We have

$$\begin{aligned}
& \sum_j \left\langle V_1(z_1) \dots V_{j-1}(z_{j-1}) \delta_\epsilon V_j(z_j) V_{j+1}(z_{j+1}) \dots V_s(z_s) \right\rangle_\Sigma = \\
& = \sum_j \left\langle V_1(z_1) \dots V_{j-1}(z_{j-1}) \left(\oint_{C_j} s(w) V_j(z_j) \right) V_{j+1}(z_{j+1}) \dots V_s(z_s) \right\rangle_\Sigma = \\
& = \left\langle \int_{\partial(\Sigma \cup_j D_j)} s(w) V_1(z_1) \dots V_s(z_s) \right\rangle_\Sigma = \left\langle \int_{\Sigma \cup_j D_j} ds(w) V_1(z_1) \dots V_s(z_s) \right\rangle_\Sigma = 0.
\end{aligned} \tag{3.294}$$

The vanishing of the sum in the first line is the Ward identity of local (i.e. gauged) 10d supersymmetry. We stress again that, in order for the Ward identity to hold, it is necessary that $s(w)$ is a global 1-form. $Q_\alpha(w)$ has weight $h = 1$ hence $s(w')' = s(w)$ for all holomorphic change of coordinates $w \mapsto w' = w'(w)$. Then $s(w)$ is globally well-defined iff it has no branch-cuts, i.e. *iff* $Q_\alpha(w)$ is *mutually-local with respect to all inserted vertices* $V_j(z_j)$. By the analysis of Sect. 3.1, this condition is equivalent to the GSO projection. Hence

The SUSY Ward identities hold if and only if we impose the GSO projection

Tadpoles A special case is the v.e.v. of a massless vertex $\langle V(z) \rangle$ called a *tadpole amplitude*. To be non-zero, $V(z)$ should be bosonic. Then it may be written as

$$V = \delta_\epsilon V_F, \tag{3.295}$$

for a suitable fermionic vertex V_F . Hence

$$\langle V \rangle_\Sigma = \langle \delta_\epsilon V_F \rangle_\Sigma = 0. \tag{3.296}$$

Thus *after the GSO projection, all tadpoles vanish*. The vanishing of the tadpoles is related to the stability of the vacuum. We expect that a vacuum with unbroken SUSY is automatically stable. The above results are consistent with this physical expectation.

We emphasize that the world-sheet conserved currents associated with spacetime supersymmetry are either purely left-moving or purely right-moving. Thus in the oriented closed string we may get supercharges from both the left and the right side, and only if on that side we have imposed the GSO projection. We have two inequivalent constructions: the GSO projections on the two sides may be on the same chirality or on opposite chirality. We shall recover these results from a different perspective in Chap. 5.

3.9 Open Strings: Chan–Paton Degrees of Freedom

The open string has two endpoints. Extended quantum systems usually have degrees of freedom which live on their boundaries, in addition to the ones propagating on the bulk. The open string is no exception. At each end of the open string, we may add a new quantum degree of freedom—called the *Chan–Paton* (CP) *label*—with finitely many states, that is, whose Hilbert space is $\mathbb{C}^n$ for some $n \in \mathbb{N}$. The on-shell open string states then have the form

$$|O; k; ij\rangle \quad i, j = 1, 2, \dots, n, \quad (3.297)$$

where $\{O e^{ik \cdot X}(x)\}$ is a BRST-invariant vertex whose insertion on the boundary creates a BRST-invariant state of the string quantized in the strip. In Eq. (3.297) i and j label the Chan–Paton state of the left and right endpoints, respectively. The world-sheet energy–momentum tensor (and hence the BRST charge Q) are the same as before, with no dependence on the boundary degrees of freedom: BRST invariance and the no-ghost theorem therefore work as in the previous sections. Spacetime Poincaré symmetry is also preserved since the new d.o.f. are inert under it. We shall check below that the addition of CP d.o.f. to the open bosonic string also preserves unitarity. The situation with the superstring is trickier, and shall be discussed in Sects. 5.6 and 10.1.

At the birth of string theory, when it was seen as a phenomenological model of strong interactions (see Sect. 1.1.1), the addition of Chan–Paton labels was quite natural. They stood for the quarks’ flavor quantum numbers: an open string was interpreted as a quark–antiquark pair connected by a color flux-tube. From our present fundamental perspective, we consider the addition of Chan–Paton d.o.f. as a way to construct *more general* consistent string theories. Later in the book,³⁵ we shall give a deeper physical interpretation of this new quantum number; for the moment we content ourselves with the original “naive” idea borrowed from the quark model of hadrons.

From Eq. (3.297) we see that in the open string sector now we have n^2 tachyons, n^2 massless vectors per each transverse polarization, etc. To describe these states, it is convenient to introduce a basis $\{\lambda_a\}$ ($a = 1, \dots, n^2$) of the $\mathbb{R}$ -space of $n \times n$ Hermitian matrices orthonormal with respect to the trace inner product

$$\text{tr}(\lambda^a \lambda^b) = \delta^{ab} \quad a, b = 1, \dots, n^2. \quad (3.298)$$

The λ^a ’s yield a complete set of states for the CP d.o.f. on the two endpoints. We shall always write the open string states in the λ^a -basis

$$|O; k; a\rangle \equiv \sum_{i,j=1}^n |O; k; ji\rangle \lambda_{ij}^a. \quad (3.299)$$

³⁵ See Chap. 6.

The corresponding BRST-invariant vertex then becomes

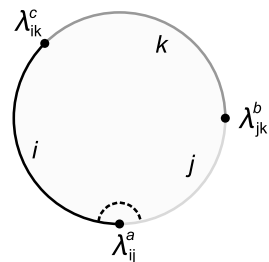
$$c(x)\lambda_{ij}^a \mathcal{O}(x) e^{ik \cdot X(x)}. \quad (3.300)$$

The matrices $\{\lambda^a\}$ generate the Lie algebra $\mathfrak{u}(n)$ in its defining representation. Hence all states of the open string transform in the adjoint representation of a $U(n)$ Chan–Paton symmetry which acts on the Chan–Paton labels in the obvious way. In the pre-historical interpretation of the open strings as mesons, this Chan–Paton symmetry was just the flavor symmetry of hadrons. In particular, the n^2 massless vectors transform in the adjoint of an $U(n)$ Chan–Paton symmetry. In a consistent theory the massless vectors can be charged under an internal symmetry only if this symmetry is *gauged* and the vectors are the gauge fields. The charged massless vectors should form precisely *one copy* of the adjoint representation of the gauge group. Therefore the $U(n)$ Chan–Paton symmetry should be promoted to an $U(n) = U(1) \times SU(n)$ gauge symmetry in physical spacetime, gauged by the open string massless vectors. If the theory is consistent (as we know it is), at low energy these vectors should have the usual Yang–Mills interactions, since these interactions are *universal* for soft “colored” massless vectors (universality of gauge interactions [30–32]). We shall check this prediction in Sect. 4.4.

The Chan–Paton d.o.f. have trivial dynamics: their state does not evolve between the vertex operators $\lambda_{i_s j_s}^{a_s} \mathcal{O}_s e^{ik_s \cdot X}$ inserted along a boundary component in cyclic order. So the right-hand endpoint of the open string state $|O_s, k_s; i_s j_s\rangle$ must be in the same state as the left-hand endpoint of $|O_{s+1}, k_{s+1}; i_{s+1} j_{s+1}\rangle$, i.e. $j_s \equiv i_{s+1}$. The CP d.o.f. associates a label $i \in \{1, 2, \dots, n\}$ with each connected *boundary arc* between two vertex insertions on the boundary $\partial\Sigma$ of the world-sheet, as well as on each connected component of $\partial\Sigma$ without insertions. See Fig. 3.1 for the example of a disk with three boundary insertions. In computing path integrals, we sum over the CP labels on each arc. Hence a boundary component S^1 with k vertex insertions, cyclically ordered along S^1 , multiplies the amplitude by the CP factor

$$\sum_{j_1, j_2, \dots, j_k=1}^n \lambda_{j_k j_1}^{a_1} \lambda_{j_1 j_2}^{a_2} \lambda_{j_2 j_3}^{a_3} \dots \lambda_{j_{k-1} j_k}^{a_k} = \text{tr}(\lambda^{a_1} \lambda^{a_2} \dots \lambda^{a_k}). \quad (3.301)$$

Fig. 3.1 A 3-open-string-disk amplitude. The open string vertices split the boundary in 3 arcs drawn in different colors. The b.c. in each arc is specified by its CP label. The open string state λ_{ij}^a (dashed curve) has the 1st (resp. 2nd) end in the i (resp. j) state



To get the full S -matrix we have to sum over the distinct cyclic orders along the boundaries.

Unitarity We show that adding the CP d.o.f. is consistent with unitarity. It is enough to check that cutting the free propagator of the oriented open string —i.e. the amplitude on the strip $\Sigma = \{(\sigma, \tau) : 0 \leq \sigma \leq \pi, \tau \in \mathbb{R}\}$ —and inserting a complete set of *physical states* reproduces the free propagation. The physical amplitude is a product of an amplitude for the CP d.o.f. and one for the bulk d.o.f. Assuming that the last one satisfies the bulk cutting relation $\sum_O |O\rangle\langle O^\dagger| = \text{Id}_{\text{bulk}}$, we get

$$\sum_{a,O} |O; ij\rangle \lambda_{ji}^a \lambda_{kl}^a \langle O^\dagger; lk| = \delta_{il} \delta_{jk} \text{Id}_{\text{bulk}}, \quad (3.302)$$

and unitarity requires the oriented open string CP gauge group to be $U(n)$.

Non-Oriented Strings

We sketch the construction of non-oriented open string theories with CP d.o.f. To make the story shorter we consider the bosonic string, leaving the obvious extension to the superstring to the reader. We write (σ, τ) , $0 \leq \sigma \leq \pi$, $-\infty < \tau < \infty$, for the coordinates on the strip and Ω for the orientation-reversing world-sheet parity operation.

In the open bosonic string the unitary operator Ω is defined by the property

$$\Omega X^\mu(\sigma, \tau) \Omega^{-1} = X^\mu(\pi - \sigma, \tau). \quad (3.303)$$

In view of the open string mode expansion

$$X^\mu(\sigma, \tau) = 2\alpha' p^\mu \tau + i\sqrt{2\alpha'} \sum_{n \neq 0} \frac{1}{n} \alpha_n^\mu e^{-in\tau} \cos(n\sigma), \quad (3.304)$$

Eq. (3.303) yields

$$\Omega \alpha_n^\mu \Omega^{-1} = (-1)^n \alpha_n^\mu, \quad n \in \mathbb{Z}. \quad (3.305)$$

States/operators may be classified according to their parity eigenvalue $\omega = \pm 1$. The tachyon vertices are Ω -even in both the closed and open strings. Comparing Eq. (3.305) with the mass formula $1 + \alpha' m^2 = N$, we see that for on-shell states (in the absence of CP d.o.f.)

$$\Omega|\psi\rangle = \omega_\psi |\psi\rangle, \quad \text{with} \quad \omega_\psi = (-1)^{1+\alpha' m_\psi^2}. \quad (3.306)$$

World-sheet parity is multiplicatively conserved in interactions.

The Ω Projection Given a consistent *oriented* string theory, we can make a new *non-oriented* string theory by restricting to the states with $\omega = +1$. Before introducing Chan–Paton d.o.f., in the open string sector the states with odd $\alpha' m^2$ remain and the

ones with even $\alpha' m^2$ (including the photon) are projected out. The multiplicative conservation of ω guarantees that unitarity survives the projection.

An important issue in the non-oriented theory is the treatment of the Chan–Paton d.o.f. Since they live on the endpoints of the open string, world-sheet parity reverses their order

$$\Omega|\psi; ij\rangle = \omega_\psi |\psi; ji\rangle. \quad (3.307)$$

Ω is a symmetry of the oriented theory; to get the unoriented theory we must restrict to states invariant under Ω . We choose our basis $\{\lambda_{ij}^a\}$ of $n \times n$ matrices to consist of elements of definite parity, i.e.

$$\lambda_{ji}^a = s^a \lambda_{ij}^a \quad \text{with} \quad s^a \in \{\pm 1\}. \quad (3.308)$$

The eigenvalue of Ω when acting on the states (3.299) is then $\omega = \omega_\psi s^a$

$$\Omega|\psi; a\rangle = \omega_\psi s^a |\psi; a\rangle, \quad (3.309)$$

and hence the states surviving in the non-oriented open string are

$$\begin{aligned} &\bullet \quad \alpha' m^2 \text{ even} \quad \lambda^a \text{ antisymmetric} \\ &\bullet \quad \alpha' m^2 \text{ odd} \quad \lambda^a \text{ symmetric.} \end{aligned} \quad (3.310)$$

For the massless gauge bosons, the Chan–Paton factor is an *antisymmetric* $n \times n$ matrix, i.e. the gauge group is $SO(n)$. The states at even mass levels transform in the *adjoint*, and those at the odd mass level in the traceless symmetric plus the trivial representation.

More General Constructions

The oriented theory has other orientation-reversing symmetries, obtained by combining Ω with a $U(n)$ rotation γ

$$\Omega_\gamma |\psi; ij\rangle \stackrel{\text{def}}{=} \omega_\psi \gamma_{jj'} |\psi; j'i'\rangle \gamma_{i'i}^{-1}. \quad (3.311)$$

Therefore we may form more general unoriented theories by projecting on $\Omega_\gamma = +1$ with $\gamma \in U(n)$. This procedure is consistent with the interactions for the same reason as for the original Ω -projection. One has

$$\Omega_\gamma^2 |\psi; ij\rangle = [(\gamma^t)^{-1} \gamma]_{ii'} |\psi; i'j'\rangle [\gamma^{-1} \gamma^t]_{j'j}. \quad (3.312)$$

We claim that consistency requires $\Omega_\gamma^2 = +1$: indeed suppose $\Omega_\gamma^2 \neq +1$ while the Ω_γ -projection yields a consistent non-oriented string theory. Then the projection on the $+1$ eigenspace of Ω_γ^2 yields a consistent *oriented* open string theory whose gauge group is the commutant of $\gamma^{-1} \gamma^t$ in $U(n)$. But the only consistent gauge group in

the oriented open theory is $U(n)$ —cf. (3.302)—hence $\gamma^{-1}\gamma'$ must be proportional to the identity by Schur's lemma, and then $\Omega_\gamma^2 = 1$. This implies³⁶

$$\gamma^t = \pm\gamma, \quad (3.313)$$

i.e. γ is either symmetric or antisymmetric. A general change of the Chan–Paton basis $\lambda^a \rightarrow U\lambda^a U^{-1}$ transforms γ as

$$\gamma \rightarrow U^t \gamma U. \quad (3.314)$$

If γ is symmetric, we may find a basis such that $\gamma = 1$, and we get back the Ω -projection considered before. If γ is antisymmetric there is a basis where [33]

$$\gamma = M \equiv \begin{pmatrix} 0 & \mathbf{1}_k \\ -\mathbf{1}_k & 0 \end{pmatrix}, \quad (3.315)$$

where $\mathbf{1}_k$ is the $k \times k$ identity matrix and $n = 2k$. Note that $M^{-1} \equiv -M \equiv M^t$. We choose a basis for the Chan–Paton wave-functions such that

$$M^{-1}(\lambda^a)^t M = \epsilon^a \lambda^a, \quad \epsilon^a = \pm 1. \quad (3.316)$$

The world-sheet parity eigenvalues now is

$$\Omega_\gamma |\psi, a\rangle = \omega_\psi \epsilon^a |\psi, a\rangle, \quad (3.317)$$

and the non-oriented open string spectrum is

$$\begin{aligned} \bullet \quad \alpha' m^2 \text{ even} \quad & M^{-1}(\lambda^a)^t M = -\lambda^a \\ \bullet \quad \alpha' m^2 \text{ odd} \quad & M^{-1}(\lambda^a)^t M = +\lambda^a. \end{aligned} \quad (3.318)$$

At the even mass levels, including the gauge bosons, we get the adjoint of the symplectic group $Sp(k)$. We conclude that in the non-oriented open bosonic string the gauge group is either $SO(n)$ or $Sp(k)$ (a.k.a. $USp(2k)$).

Appendix: Details on the No-Ghost Theorem

In this appendix we show that the OCQ physical Hilbert space $\mathbf{H}$ of the superstring is positive-definite. We focus on the left-moving sector of the closed string; the analysis of the open string is the same up to the usual factors of 2 in the center-of-mass momentum. We set $\alpha' = 2$.

³⁶ **Proof:** $\gamma^{-1}\gamma^t$ belongs to the center $U(1)$ of $U(n)$, i.e. $\gamma^t = e^{i\theta}\gamma$. Taking the transpose of this equality we get $e^{i\theta} = \pm 1$.

Our strategy is to construct an isomorphism of $\mathbf{H}$ with the light-cone Hilbert space $H_{\text{l.c.}}$ which is manifestly positive. The proof works only in critical dimension $d = 10$, as expected.

No-Ghosts: the First Proof

Since p_μ and $(-1)^F$ (anti)commute with L_m and G_r , the construction of OCQ physical/null states, and hence of the OCQ Hilbert space $\mathbf{H} = H_{\text{phys}}/H_{\text{null}}$, may be performed independently in each momentum and $(-1)^F$ eigenspace. We prove the isomorphism in a $p^\mu \neq 0$ eigenspace. At zero-momentum the light-cone gauge makes no sense, and this leads to the peculiar phenomena discussed in the main text. Since the light-cone story is not manifestly Lorentz covariant, the explicit form of the isomorphism $\mathbf{H} \simeq H_{\text{l.c.}}$ will depend on the Lorentz frame, and gets simpler by a clever choice of it. Inside the big *indefinite* Hilbert space $H \equiv H_{\text{SCFT}}$ of the (left-moving) matter SCFT,³⁷ we consider the subspace

$$\mathcal{H} = \left\{ |\psi\rangle \in H \mid p^+|\psi\rangle = 1, p^i|\psi\rangle = 0 \right\} \quad (3.319)$$

($i = 1, 2, \dots, d-2$ is the transverse vector index). The subspace of *on-shell* states in $\mathcal{H}$ is

$$\mathcal{H}_{\text{on-shell}} = \left\{ |\psi\rangle \in H \mid p^+|\psi\rangle = 1, p^i|\psi\rangle = 0, p^-|\psi\rangle = (N - \nu)|\psi\rangle \right\}, \quad \nu = \begin{cases} 0 & \text{R} \\ \frac{1}{2} & \text{NS} \end{cases} \quad (3.320)$$

where N is the “naive” level operator.³⁸ Any $p_\mu \neq 0$ state may be mapped in an element of $\mathcal{H}$ by a suitable Lorentz rotation, so the restriction to $\mathcal{H}$ yields no loss generality. A state $|\psi\rangle \in \mathcal{H}$ is physical (in the OCQ sense) iff it satisfies the conditions (3.58).

Lemma 3.3 *An R -sector state $|\psi\rangle \in \mathcal{H}$ with $G_0|\psi\rangle = 0$ has the form $G_0|\psi'\rangle$ for $|\psi'\rangle \in \mathcal{H}_{\text{on-shell}}$.*

Proof (Cf. [6] Sect. 2) Write $G_0 = G_0^0 + G_0^1$ where

$$G_0^0 = -\oint \frac{dz}{2\pi i} z^{1/2} \psi(z)^0 \partial X^0(z), \quad G_0^1 = \sum_{k=1}^{d-1} \oint \frac{dz}{2\pi i} z^{1/2} \psi(z)^k \partial X^k(z) \quad (3.321)$$

and $G_0^+ = -G_0^0 + G_0^1$. One has

$$\{G_0, G_0^+\} = 2(p_0^2 + \mathbf{p}^2 + \text{absolute level operator}) \equiv M, \quad (3.322)$$

with M a strictly positive operator since $p_\mu \neq 0$ in $\mathcal{H}$. One has $|\psi\rangle = G_0^+ M^{-1} |\psi\rangle$. $\square$

Definition 3.1 $|\psi\rangle \in \mathcal{H}$ is a *transverse state* if it satisfies the conditions

$$L_n |\psi\rangle = G_r |\psi\rangle = \alpha_n^+ |\psi\rangle = \psi_r^+ |\psi\rangle = 0 \quad \text{for } n, r > 0. \quad (3.323)$$

We write $\mathcal{T} \subset \mathcal{H}$ for the subspace of transverse states. Any state $|\psi\rangle \in \mathcal{H}$ may be written as

$$|\psi\rangle = P_\psi |0; p^-\rangle, \quad (3.324)$$

where P_ψ is in the algebra $\mathfrak{M}_-$ generated by the negative modes α_{-n}^μ , ψ_{-r}^μ , and $|0; p^-\rangle$ is an oscillator ground state with momentum as in (3.319). The last two conditions in (3.323) say that if $|\psi\rangle \in \mathcal{T}$ then P_ψ does not contain any ψ_{-r}^- , α_{-n}^- oscillator. This suffices to conclude that *the inner product is positive semi-definite in $\mathcal{T}$* : indeed states of the form $\alpha_{-n}^+ |\chi\rangle$ or $\psi_{-r}^+ |\chi\rangle$ are orthogonal to all states in $\mathcal{T}$ (including themselves), while states without α_{-n}^+ , ψ_{-r}^+ oscillators have positive norm.

³⁷ We ignore the FP ghosts, which are frozen in their “vacuum”. **Beware:** confusingly the term “ghosts” is used in string theory in two very different senses: Faddeev-Popov ghost *fields* versus negative-norm *states*. The no-ghost theorem refers to the absence of ghosts in the second sense.

³⁸ N is not positive-definite since it counts the time-like oscillator α_{-n}^0 , ψ_{-r}^0 with a minus sign.

Lemma 3.4 Let $\{|t\rangle\}$ be a basis of $\mathcal{T}$ with elements of definite level N_t . A basis of $\mathcal{H}$ is given by³⁹

$$| \{a_n, b_r, c_n, d_r\}, t \rangle \equiv \left(\prod_{r>0} \vec{G}_{-r}^{a_r} \right) \left(\prod_{n>0} \vec{L}_{-n}^{b_n} \right) \left(\prod_{r>0} \vec{\psi}_{-r}^{+c_r} \right) \left(\prod_{n>0} \vec{\alpha}_{-n}^{+d_n} \right) |t\rangle, \quad (3.325)$$

where $a_r, c_r = 0, 1, b_n, d_n = 0, 1, 2, 3, \dots$. The state (3.325) has level

$$N = \sum_{r>0} r(a_r + c_r) + \sum_{n>0} n(b_n + d_n) + N_t, \quad (3.326)$$

and

$$\langle t|t'\rangle = 0 \Rightarrow \langle \{a_n, b_r, c_n, d_r\}, t | \{a'_n, b'_r, c'_n, d'_r\}, t'\rangle = 0. \quad (3.327)$$

Proof The first statement follows from $L_m = \alpha_m^- + \dots$, $G_r = \psi_r^- + \dots$ and the Poincaré–Birkhoff–Witt (PBW) theorem [21]. The second one is obvious. For the third one, write the LHS of (3.327) as $\langle t|P_1 P_2|t'\rangle$, where P_1 is a product of L_n 's, G_r 's, α_n^- 's, and ψ_r^- 's with positive indices, while P_2 is a product of the corresponding operators with negative indices. Using again PBW, expand $P_1 P_2$ in the basis given by monomials in the L_n 's, G_r 's, α_n^- , ψ_r^- ordered in increasing level; the only monomials which do not vanish when sandwiched between $\langle t|, |t'\rangle$ are constants and L_0 . $\square$

Note that N and p^- are diagonal in the basis (3.325). Thus a basis of the on-shell Hilbert space $\mathcal{H}_{\text{on-shell}}$ is just given by the vectors in (3.325) with $p^- = N - \nu$. Then **Lemma 3.4** implies

$$\mathcal{H} = \mathcal{V} \oplus \mathcal{S}, \quad (3.328)$$

where $\mathcal{V}$ is the subspace spanned by the vectors in (3.325) with

$$s \equiv \sum_{r>0} r a_r + \sum_{n>0} n b_n = 0, \quad (3.329)$$

and $\mathcal{S}$ the subspace spanned by the on-shell basis vectors with $s > 0$. The elements of $\mathcal{S}$ are spurious by construction and conversely all spurious states are in $\mathcal{S}$.

Lemma 3.5 Assume $d = 10$. Then

NS sector the operators $G_{1/2}$ and $\check{G}_{3/2} \equiv G_{3/2} + 2L_1 G_{1/2}$ map each subspace $\mathcal{S}, \mathcal{V}$ into itself;
R-sector G_0 and $L_1 G_0$ map each subspace $\mathcal{S}, \mathcal{V}$ into itself.

Proof For $\mathcal{V}$ it is clear. We need to check the action of the operators on $\mathcal{S}$.

NS sector: a NS state in $\mathcal{S}$ may be written in the form $G_{-1/2}|A\rangle + G_{-3/2}|B\rangle$ with $L_0|A\rangle = 0$ and $L_0|B\rangle = -|B\rangle$. Using Eqs. (2.433) and (2.434) we get

$$G_{1/2}(G_{-1/2}|A\rangle + G_{-3/2}|B\rangle) = -G_{-1/2}G_{1/2}|A\rangle + G_{-3/2}G_{1/2}|B\rangle + 2L_{-1}|B\rangle \in \mathcal{S} \quad (3.330)$$

$$G_{3/2}(G_{-1/2}|A\rangle + G_{-3/2}|B\rangle) = 2L_1|A\rangle + (-2 + d)|B\rangle \mod \mathcal{S} \quad (3.331)$$

$$L_1 G_{1/2}(G_{-1/2}|A\rangle + G_{-3/2}|B\rangle) = -L_1|A\rangle - 4|B\rangle \mod \mathcal{S}. \quad (3.332)$$

R-sector: one writes a state in $\mathcal{S}$ in the general form $L_{-1}|A\rangle + G_{-1}|B\rangle$. Then

³⁹ The symbol $\prod$ means that the operators are in increasing level order.

$$G_0(L_{-1}|A\rangle + G_{-1}|B\rangle) = L_{-1}G_0|A\rangle + \frac{1}{2}G_{-1}|A\rangle - G_{-1}G_0|B\rangle + 2L_{-1}|B\rangle \in \mathcal{S} \quad (3.333)$$

$$L_1G_0L_{-1}|A\rangle = -G_{-1}L_1G_0|A\rangle + 2L_{-1}L_1|A\rangle + (4L_0 - \frac{3}{2}G_0^2)|A\rangle \quad (3.334)$$

$$L_1G_0L_{-1}|B\rangle = L_{-1}L_1G_0|B\rangle + \frac{1}{2}G_{-1}L_1|B\rangle + 3(L_0 + \frac{3}{8})G_0|B\rangle, \quad (3.335)$$

while the on-shell condition yield

$$G_0^2|A\rangle = -|A\rangle, \quad L_0|A\rangle = \frac{d-16}{16}|A\rangle, \quad L_0|B\rangle = \frac{d-16}{16}|B\rangle, \quad (3.336)$$

so that for $d = 10$ the RHS of (3.334), (3.335) belong to $\mathcal{S}$. $\square$

Theorem 3.3 (No-Ghost Theorem) *In the OCQ setup, in $d = 10$ with the mass-shell conditions*

$$(L_0 - \frac{1}{2})|\text{phys}\rangle_{\text{NS}} = 0, \quad G_0|\text{phys}\rangle_{\text{R}} = 0, \quad (3.337)$$

the physical Hilbert space $\mathbf{H} \equiv H_{\text{phys}}/H_{\text{null}}$ is positive-definite.

Proof We follow [6]. From (3.328) a physical state $|\text{phys}\rangle \in \mathcal{H}$ is written in a unique way as

$$|\text{phys}\rangle = |v\rangle + |s\rangle, \quad v \in \mathcal{V}, \quad s \in \mathcal{S}. \quad (3.338)$$

We claim that $|v\rangle$ and $|s\rangle$ are separately physical. For the NS sector this is obvious from (3.58) and **Lemma 3.5**. For the R-sector **Lemma 3.3** gives

$$|\text{phys}\rangle \in G_0\mathcal{V} \oplus G_0\mathcal{S}, \quad (3.339)$$

while $L_1G_0\mathcal{V} \subseteq \mathcal{V}$ and $L_1G_0\mathcal{S} \subset \mathcal{S}$ by **Lemma 3.5** and the claim follows from (3.58). Then $|s\rangle$ is both physical and spurious, hence null, and, modulo null states, all physical states belong to $\mathcal{V}_{\text{on-shell}} \subset \mathcal{V}$. But a physical state in $\mathcal{V}$ is necessarily transverse: indeed $\mathcal{V}$ consists of states with

$$\alpha_n^+|v\rangle = \psi_r^+|v\rangle = 0 \quad n, r > 0, \quad (3.340)$$

and a vector which satisfies (3.340) together with the physical conditions $L_n|v\rangle = G_r|v\rangle = 0$ ($n, r > 0$) is transverse by definition; cf. Eq. (3.323). We conclude that *all physical states are transverse modulo null states*. That is, modulo null states a physical state belongs to the space $\mathcal{T}_{\text{on-shell}} \subset \mathcal{T}$. We already know that $\mathcal{T}_{\text{on-shell}}$ has a semi-definite inner product. Let us show that the Hilbert space $\mathcal{T}_{\text{on-shell}}$ is actually positive-definite. A norm zero element $|\omega\rangle \in \mathcal{T}_{\text{on-shell}}$ is automatically orthogonal to all elements of $\mathcal{T}$ and $\mathcal{S}$, and (being physical) should be null, hence spurious. Thus $|\omega\rangle \in \mathcal{T} \cap \mathcal{S}$ is the zero vector. $\square$

Additional details are obtained by writing the transverse states explicitly.

DDF States and the Spectrum-Generating Algebra

We work in the Hilbert space $\mathcal{H}_{\text{on-shell}}$. The ground states in $\mathcal{H}_{\text{on-shell}}$ at $N = 0$ are a unique tachyon state $|\text{tac}\rangle$ in the NS sector and a massless Majorana spinor $|u\rangle$ in the R one. These states are physical and non-null. $\mathcal{H}_{\text{on-shell}}$ is constructed by acting on the vacua $|\text{tac}\rangle, |u\rangle$ with the *dressed* creation operators ($n, r > 0$)

$$\hat{\alpha}_{-n}^\mu = \alpha_{-n}^\mu e^{-inx^+}, \quad \hat{\psi}_{-r}^\mu = \psi_{-r}^\mu e^{-irx^+}, \quad (\mu = +, 1, \dots, 8, -). \quad (3.341)$$

Here x^+ is the zero-mode of the world-sheet field $X^+(z)$

$$X^+(z) = x^+ + p^+z + \text{oscillators}, \quad (3.342)$$

The effect of the exponential factor e^{-inx^+} in the hatted operators (3.341) is to shift $p^- \rightarrow p^- + n$, so that they map $\mathcal{H}_{\text{on-shell}}$ to itself. Dually $|\text{tac}\rangle, |u\rangle$ are killed by the dressed annihilation operators

$$\hat{\alpha}_n^\mu = \alpha_n^\mu e^{inx^+}, \quad \hat{\psi}_r^\mu = \psi_r^\mu e^{irx^+}, \quad n, r > 0, \quad (3.343)$$

and the hatted operators satisfy the same algebra as the original operators. The lever operator N has the same form in the hatted and un-hatted operators

$$N \equiv \sum_{n \geq 1} \alpha_{-n}^\mu \alpha_{\mu n} + \sum_{r > 0} r \psi_{-r}^\mu \psi_{\mu r} = \sum_{n \geq 1} \hat{\alpha}_{-n}^\mu \hat{\alpha}_{\mu n} + \sum_{r > 0} r \hat{\psi}_{-r}^\mu \hat{\psi}_{\mu r}. \quad (3.344)$$

The basic advantage of working with hatted modes is that they (anti)commute with G_0, L_0 . For our purposes it is desirable to have operators $\mathcal{H}_{\text{on-shell}} \rightarrow \mathcal{H}_{\text{on-shell}}$ which (anti)commute with G_n, L_n for all n . Consider the “transverse” operators $\hat{\alpha}_n^i$ with $n \in \mathbb{Z} \setminus \{0\}$ and $i = 1, \dots, 8$. We work in Lorentz signature with coordinates (σ, τ) , σ periodic of period 2π , and we set $w = \sigma + \tau$

$$\hat{\alpha}_n^i = e^{inx^+} \int_0^{2\pi} \frac{dw}{2\pi} \partial X^i(w) e^{inw} = \int_0^{2\pi} \frac{dw d\theta}{2\pi} DX^i e^{inX^+} \Big|_{\text{l.c.}}, \quad (3.345)$$

where we used the light-cone gauge conditions⁴⁰ $X^+(w)|_{\text{l.c.}} = x^+ + p^+w$ and $\psi^+|_{\text{l.c.}} = 0$, while $p^+ \equiv 1$ in the Hilbert space $\mathcal{H}_{\text{on-shell}}$. This suggests to replace

$$\hat{\alpha}_n^i \rightarrow A_n^i \equiv \int_0^{2\pi} \frac{dw d\theta}{2\pi} DX^i e^{inX^+}, \quad (3.346)$$

where now $X^+(z, \theta)$ is the conformal gauge superfield. This definition is well-posed since, when acting on $\mathcal{H}_{\text{on-shell}}$, the superfields DX^i and e^{inX^+} are periodic in w of period 2π . Indeed

$$e^{inX^+(w)} = \exp\left(inx^+ + inp^+w + in \sum_{m \neq 0} \alpha_m^+ e^{-imw}\right) \equiv e^{inx^+ + inw + \text{Fourier series}}. \quad (3.347)$$

A_i has an expansion of the form

$$A_n^i = \hat{\alpha}_n^i + \text{terms containing } \alpha_m^+, \psi_r^+, (m, r \neq 0) \text{ oscillators}, \quad (3.348)$$

so, modulo *null* oscillators in the light-cone $+$ direction, the A_n^i ’s coincide with our hatted transverse operators. By construction, the A_n^i have the desired properties

$$[L_m, A_n^i] = [G_r, A_n^i] = 0, \quad [N, A_n^i] = -nA_n^i. \quad (3.349)$$

We can do a similar (but more involved) replacement for the transverse Fermi oscillators [3]

$$\hat{\psi}_r^i \rightarrow \Psi_r^i = \int \frac{dw d\theta}{2\pi} DX^i \frac{DX^+}{(\partial X^+)^{1/2}} e^{irX^+}. \quad (3.350)$$

Exercise 3.4 Check that the Fourier coefficients Ψ_r^i are well defined both in the NS and R sectors.

The inverse square-root in (3.350) looks awful, but it is a well-defined regular operator when acting on our cleverly chosen Hilbert space $\mathcal{H}_{\text{on-shell}}$

$$(\partial X^+)^{-1/2} = \sum_{\ell \geq 0} \binom{-1/2}{\ell} \left(\sum_{n \neq 0} \alpha_n^+ e^{inw} + \theta \sum_r r \psi_r^+ e^{irw} \right)^\ell. \quad (3.351)$$

⁴⁰ We suppressed the right-movers, i.e. $X(z)^+ \equiv X_L(z)^+$.

Notice that

$$\left. \frac{DX^+}{(\partial X^+)^{1/2}} \right|_{\text{light-cone}} = 0 + \theta \sqrt{p^+} \equiv \theta, \quad (3.352)$$

so that

$$\int \frac{dw d\theta}{2\pi} DX^i \frac{DX^+}{(\partial X^+)^{1/2}} e^{irX^+} \Big|_{\text{light-cone}} = \hat{\psi}_r^i, \quad (3.353)$$

in perfect analogy with Eq. (3.345). Hence,

$$\Psi_r^i = \hat{\psi}_r^i + \text{terms containing } \alpha_m^+ \text{ and } \psi_s^+, (m, s \neq 0) \text{ oscillators,} \quad (3.354)$$

and again (when acting on $\mathcal{H}_{\text{on-shell}}$)

$$[L_m, \Psi_s^i] = \{G_r, \Psi_s^i\} = 0, \quad [N, \Psi_r^i] = -r\Psi_r^i. \quad (3.355)$$

The operators A_n^i and Ψ_r^i —called the *DDF operators*⁴¹—generate, for r half-integral resp. integral, the algebras $\mathcal{A}_{\text{NS}}$ and $\mathcal{A}_{\text{R}}$ which are isomorphic to the canonical one for the original transverse oscillators, α_n^i, ψ_r^i , i.e. [34]

$$[A_m^i, A_n^j] = m \delta^{ij} \delta_{m+n,0}, \quad [A_n^i, \Psi_r^j] = 0, \quad \{\Psi_r^i, \Psi_s^j\} = \delta^{ij} \delta_{r+s,0}, \quad (3.356)$$

which is called the *spectrum-generating algebra*. As its name suggests, the algebras $\mathcal{A}_{\text{NS}}$, $\mathcal{A}_{\text{R}}$ generate the physical spectrum of the superstring, that is, the Hilbert space $\mathcal{H}_{\text{NS}}$ (resp. $\mathcal{H}_{\text{R}}$) produced by the action of $\mathcal{A}_{\text{NS}}$ (resp. $\mathcal{A}_{\text{R}}$) acting on the NS “vacuum” $|\text{tac}\rangle$ (resp. the R “vacua” $|u\rangle$) is isomorphic to the momentum eigenspace $p^+ = 1$, $p^i = 0$ of the OCQ physical space $\mathbf{H}_{\text{NS}} \equiv H_{\text{NS,phys}}/H_{\text{NS,null}}$ (resp. $\mathbf{H}_{\text{R}} \equiv H_{\text{R,phys}}/H_{\text{R,null}}$)

$$\mathcal{H}_{\text{NS,R}} \simeq \mathbf{H}_{\text{NS,R}} \Big|_{p^+=1, p^i=0}. \quad (3.357)$$

This statement is equivalent to the no-ghost theorem: from Eq. (3.356) the Fock subspaces

$$\mathcal{F}_{\text{NS}} \subset \mathbf{H}_{\text{NS}}, \quad \text{and} \quad \mathcal{F}_{\text{R}} \subset \mathbf{H}_{\text{R}}, \quad (3.358)$$

generated by acting with the DDF creation operators A_{-n}^i, Ψ_{-r}^i on the NS resp. R vacua, are isomorphic on the nose to the corresponding (physical) Hilbert spaces in the light-cone gauge $\mathcal{H}_{\text{NS,l.c.}}, \mathcal{H}_{\text{R,l.c.}}$, and hence automatically positive-definite. The states in the subspaces $\mathcal{F}_{\text{NS}}, \mathcal{F}_{\text{R}}$ are called *DDF states*; they are exactly the space of *transverse states* in the sense of **Definition 3.1**.

The operator E and the supervector $D(z, \theta)$

We introduce a quantum operator E which is diagonal in the basis (3.325) with eigenvalue

$$E = N_t - N. \quad (3.359)$$

From Eq. (3.327) it follows that eigenvectors of $N_t = N + E$ with distinct eigenvalues are orthogonal. Then N_t and E are Hermitian operators. We restate the no-ghost theorem in a convenient way:

Proposition 3.1 (Properties of E) *There exists a Hermitian operator E , acting on the Hilbert (sub)space $\mathcal{H} \subset H$ of the critical NS-R superstring, such that*

- (a) *the spectrum of E is given by $-\mathbb{N}$ in the R-sector and by $-\frac{1}{2}\mathbb{N}$ in the NS one⁴²;*
- (b) *the transverse states $|t\rangle \in \mathcal{T}$ are the zero eigenvectors of E ;*

⁴¹ After Del Giudice, Di Vecchia and Fubini [34].

⁴² $\mathbb{N}$ stands for the set of non-negative integers.

(c) E satisfies the relations

$$[E, L_n] = nL_n, \quad [E, G_r] = rG_r, \quad (3.360)$$

and hence maps physical states into physical states;

(d) the state $E|\text{phys}\rangle$ is null for all physical states $|\text{phys}\rangle$. All null states are of this form;

(e) there exist operators $D_n = (D_{-n})^\dagger$ and $B_r = (B_{-r})^\dagger$ such that

$$E = \sum_{n>0} (L_{-n}D_n + D_{-n}L_n) + \frac{1}{2} \sum_{r>0} (G_{-r}B_r + B_{-r}G_r) + \begin{cases} (D_0 - 1)(L_0 - \frac{1}{2}) & \text{NS} \\ B_0G_0 + (D_0 - 1)(L_0 - 5/8) & \text{R}; \end{cases}$$

(f) the operators B_r, D_n contain only the oscillators α_m^+ and ψ_s^+ ($m \neq 0$) and hence

$$[D_n, D_m] = [D_n, B_r] = \{B_r, B_s\} = 0 \quad (3.361)$$

$$D_n|t\rangle = G_r|t\rangle = 0 \quad D_0|t\rangle = |t\rangle \quad \forall |t\rangle \in \mathcal{T}, \quad n, r > 0; \quad (3.362)$$

(g) working on a cylinder with in/out states as $\tau \rightarrow \pm\infty$ in the subspace $\mathcal{H} \subset H$, the modes D_n, B_r may be combined in a supervector $\mathbf{D}(z, \theta)$, i.e. in a $h = -1$ superconformal pseudo-superfield

$$\mathbf{D}(z, \theta) \equiv D(z) + \theta B(z) = \sum_{n \in \mathbb{Z}} \frac{D_n}{z^{n-1}} + \theta \sum_{r \in \mathbb{Z} + \nu} \frac{B_r}{z^{r-1/2}}. \quad (3.363)$$

Then, from Eqs. (2.440)–(2.443), we have

$$[L_m, D_n] = -(2m + n)D_{m+n} \quad [L_m, B_s] = -(\frac{3}{2}m + s)B_{m+s}, \quad (3.364)$$

$$[G_r, D_n] = B_{n+r}, \quad \{G_r, B_s\} = -(3r + s)D_{r+s}; \quad (3.365)$$

(h) on the cylinder as above we have the supercurrent (i.e. chiral superfield with $h = 1/2$)

$$E(z, \theta) = \mathbf{D}(z, \theta) \mathbf{T}(z, \theta) \quad \text{so that} \quad E = \oint \frac{dz d\theta}{2\pi i} E(z, \theta). \quad (3.366)$$

Proof (a) and (b) are already known. Equation (3.360) is equivalent to **Lemma 3.5**. We stress that (3.360) is true if and only if the superstring is critical [3]. Then (c) follows from the identity

$$L_n E|\text{phys}\rangle = [L_n, E]|\text{phys}\rangle = -nL_n|\text{phys}\rangle = 0 \quad \text{for } n > 0, \quad (3.367)$$

and the similar one with $L_n \rightarrow G_r$. (d) physical states orthogonal to transverse ones are null (see proof of **theorem**), and $E|\text{phys}\rangle$ is orthogonal to all vectors in $\mathcal{T}$ since E is Hermitian and item (b) holds. Conversely, consider a basis of the space of null states whose elements have definite eigenvalue λ of E . By (b) $\lambda \neq 0$. Then $|\text{null}\rangle = E(\lambda^{-1}|\text{null}\rangle)$. (e) follows from the elementary argument that a Hermitian operator which maps physical states into physical states must have the stated form for some B_r, D_n (see, for example, [10]). In the expression of E we added terms which vanish on-shell: they do not matter in the proof of the no-ghost theorem, since a physical state is automatically on-shell. These terms are however crucial for the following statements in the **Proposition** to hold. (f) By the definition of E and Eqs. (3.349), (3.355), B_r, D_n must commute with the DDF operators A_n^i, Ψ_r^i in Eqs. (3.348), (3.354) and with the oscillators α_n^+, ψ_r^+ . Then D_0 has an expansion as a sum of monomials of the form

$$D_0 = \sum_{\{n_i\}, \{r_j\}} c(\{n_i\}, \{r_j\}) \delta(\sum_i n_i + \sum_j r_j) \prod_i \alpha_{n_i}^+ \prod_j \psi_{r_j}^+. \quad (3.368)$$

Thus D_0 acting on a transverse state is just the constant $c(0, 0)$. It follows from the explicit construction of DDF states that this constant is $1/p^+ = 1$ in our units. (g) follows from consistency

with superconformal symmetry restricted to the cylinder with our in/out states. The two components of $\mathbf{D}(z, \theta)$ are, respectively, a Killing vector $D(z)$ and a Killing spinor $B(z)$. Let us recall a few geometrical facts: Killing vector/spinors may exist only for the torus with no puncture and the sphere with at most two punctures. In the absence of punctures the Killing vector/spinor forms a finite-dimensional space, and hence cannot form a non-trivial superfield. We are left with the sphere with two punctures (< 2 punctures being special cases) at which we have to specify suitable boundary conditions. This shows that $\mathbf{D}(z, \theta)$ cannot be defined in general in the SCFT, but only in our particular setup. Finally **(h)** also follows. $\square$

Note 3.8 Working in the subspace $\mathcal{H}$, we may replace the oscillator algebra generated by α_n^μ, ψ_r^μ ($\mu = \pm, 1, \dots, 8$), with the algebra $\mathcal{A}$ generated by the operators

$$A_n^i, L_n, D_n, \Psi_r^i, G_r, B_r. \quad (3.369)$$

These operators have a complicated expression in terms of the original oscillators, but they satisfy a simple superalgebra: see Eqs. (2.432)–(2.434), (3.349), (3.355), (3.356), (3.361), (3.364), (3.365), together with

$$[A_n^i, D_m] = [A_n^i, B_s] = [\Psi_r^i, D_m] = \{\Psi_r^i, B_s\} = 0. \quad (3.370)$$

References

1. Kosteletzky, V.A., Lechtenfeld, O., Lerche, W., Samuel, S., Watamura, S.: Conformal techniques, bosonization and tree level string amplitudes. Nucl. Phys. B **288**, 173–232 (1987)
2. Bröcker, T., tom Dieck, T.: Representations of compact Lie groups. Springer (1985)
3. Scherk, J.: An introduction to the theory of dual models and strings. Rev. Mod. Phys. **47**, 123–164 (1975)
4. Gupta, S.: Theory of longitudinal photons in quantum electrodynamics. Proc. Phys. Soc. **63A**, 681–691 (1950)
5. Brink, L., Olive, D.I.: The physical state projection operator in dual resonance models for the critical dimension of space-time. Nucl. Phys. B **56**, 253–265 (1973)
6. Corrigan, E., Goddard, P.: The absence of ghosts in the dual fermion model. Nucl. Phys. B **68**, 189–202 (1974)
7. Green, M.B., Schwarz, J.H.: Covariant description of superstrings. Phys. Lett. B **136**, 367 (1984)
8. Berkovits, N.: Super-poincare covariant quantization of the superstring. JHEP **04**, 018 (2000). [arXiv:hep-th/0001035](https://arxiv.org/abs/hep-th/0001035)
9. Berkovits, N.: ICTP Lectures on covariant quantization of the superstring. hep-th/0209059
10. Green, M.B., Schwarz, J., Witten, E.: Superstring Theory, vols. 1, 2. Cambridge University Press, Cambridge (1987)
11. Becker, K., Becker, M., Schwarz, J.: String Theory and M-Theory. Cambridge University Press, Cambridge (2007)
12. Berkovits, N., Gomez, H.: An introduction to pure spinor superstring theory. [arXiv:1711.09966](https://arxiv.org/abs/1711.09966)
13. Becchi, C., Rouet, A., Stora, R.: Renormalization of gauge theories. Ann. Phys. **98**, 287–321 (1976)
14. Tyutin, I.V.: Gauge Invariance in Field Theory and Statistical Physics in Operator Formalism, Lebedev Physics Institute preprint 39 (1975). [arXiv:0812.0580](https://arxiv.org/abs/0812.0580)
15. Kugo, T., Uehara, S.: General procedure of gauge fixing based on BRS invariance principle. Nucl. Phys. B **197**, 378 (1982)
16. Becchi, C.: Introduction to BRS symmetry. [arXiv:hep-th/9607181](https://arxiv.org/abs/hep-th/9607181)
17. Weibel, C.A.: An Introduction to Homological Algebra. Cambridge University Press (1994)
18. Henneaux, M.: Remarks on the cohomology of the BRS operator in string theory. Phys. Lett. B **177**, 35–38 (1986)

19. Henneaux, M.: BRST cohomology of the fermionic string. *Phys. Lett. B* **183**, 59–64 (1987)
20. Freeman, M.D., Olive, D.I.: BRS cohomology in string theory and the no-ghost theorem. *Phys. Lett. B* **175**, 151–154 (1986)
21. Bourbaki, N.: *Éléments de mathématique. Groupes et algèbres de Lie. Chapitre 1: Algèbres de Lie*. Springer (2007)
22. Friedan, D., Martinec, E., Shenker, S.: Conformal invariance, supersymmetry, and string theory. *Nucl. Phys. B* **271**, 93 (1986)
23. Ohta, N.: BRST cohomology in superstring theories. *Phys. Lett. B* **179**, 347–351 (1986)
24. Furuuchi, K., Ohta, N.: On the no-ghost theorem in string theory. *Prog. Theor. Phys.* **116**, 601–604 (2006). [arXiv:hep-th/0607105](https://arxiv.org/abs/hep-th/0607105)
25. Dedushenko, M.: Superstring BRST no-ghost theorem at arbitrary picture number. [arXiv:1212.3592](https://arxiv.org/abs/1212.3592)
26. Berkovits, N., Zwiebach, B.: On the picture dependence of Ramond-Ramond cohomology. *Nucl. Phys. B* **523**, 311–343 (1998). [arXiv:hep-th/9711087](https://arxiv.org/abs/hep-th/9711087)
27. Witten, E.: Interacting field theory of open superstrings. *Nucl. Phys. B* **276**, 291–324 (1986)
28. Witten, E.: Notes on super-Riemann surfaces and their moduli. *Pure Appl. Math. Quart.* **15**(1), 57–211 (2019). [arXiv:1209.2459](https://arxiv.org/abs/1209.2459)
29. Deser, S., Zumino, B.: Consistent supergravity. *Phys. Lett. B* **62**, 335 (1976)
30. Weinberg, S.: Derivation of gauge invariance and the equivalence principle from Lorentz invariance of the S-matrix. *Phys. Lett.* **9**, 357–359 (1964)
31. Weinberg, S.: Photons and gravitons in S-matrix theory: Derivation of charge conservation and equality of gravitational and inertial mass. *Phys. Rev.* **135**, B1049–B1056 (1964)
32. Weinberg, S.: Photons and gravitons in perturbation theory: derivation of Maxwell's and Einstein's equations. *Phys. Rev.* **138**, B98–B1002 (1965)
33. Lang, S.: *Algebra*. Springer (2002)
34. Del Giudice, E., Di Vecchia, P., Fubini, S.: General properties of the dual resonance model. *Ann. Phys.* **70**, 378–398 (1972)

Chapter 4

Bosonic String Amplitudes



Abstract We compute quantum amplitudes in the various bosonic string theories (open, closed, oriented, and non-oriented) with particular emphasis on world-sheets Σ with $\chi(\Sigma) \geq 0$ which yield the leading perturbative contributions. We construct and study the physical properties of the Veneziano and Shapiro–Virasoro amplitudes. We start by reviewing the techniques to compute path integrals for 2d non-compact free scalars and Fermi b, c systems living on general 2-manifolds, giving detailed expressions for all surfaces with non-negative Euler characteristic.

The perturbative expansion in (super)string theory is a sum of contributions from connected (super-)world-sheets Σ of all possible topologies. Each term in the sum is the integral over the (super)moduli space of a (S)CFT amplitude computed by a path integral over the 2d matter and ghost fields propagating on Σ ; cf. Sect. 1.4. The amplitude from surfaces of a given topology scales with the string coupling $g \equiv e^\Phi$ as $g^{-\chi(\Sigma)}$. We are particularly interested in the leading contributions as $g \rightarrow 0$. There are three world-sheet topologies with $\chi(\Sigma) > 0$ which give the “tree-level” contributions:

- sphere S^2 , $\chi(S^2) = 2$
- disk D , $\chi(D) = 1$
- projective $\mathbb{R}$ -plane $\mathbb{RP}^2$, $\chi(\mathbb{RP}^2) = 1$;

cf. **BOX 1.5**. For oriented closed strings, the sphere is the only “tree-level” term. For open oriented string, we also have the disk D (conformally equivalent to the upper half-plane $\mathcal{H}$). For non-oriented strings in addition we have the $\mathbb{RP}^2$ contribution.

At one-loop, i.e. $\chi = 0$, we have four topologies:

- torus $T^2 = S^1 \times S^1$
- Klein bottle (non-oriented strings)
- annulus (open strings)
- Möbius band (non-oriented open strings).

For a string moving in flat spacetime, the world-sheet QFT is free, and the path integral is a product of path integrals over each decoupled free field. We consider the path integrals over each field one by one.

4.1 Path Integrals for Non-compact Scalars

Let Σ be a Riemannian 2-manifold; we write its metric in the local form

$$ds^2 = g(z, \bar{z}) dz d\bar{z} \quad \text{so} \quad \text{Vol}(\Sigma) = \int_{\Sigma} g(z, \bar{z}) d^2z, \quad (4.1)$$

where d^2z stands for $i dz \wedge d\bar{z} \equiv 2 dx dy$. We consider the path integral over the *d non-compact* scalar fields $X^\mu(z, \bar{z})$ ($\mu = 0, 1, \dots, d-1$) living on the 2d manifold Σ , in the presence of an arbitrary source $J(z, \bar{z})_\mu$

$$Z[J] = \int [dX] \exp \left\{ -\frac{1}{2\pi\alpha'} \int_{\Sigma} d^2z \partial X^\mu \bar{\partial} X_\mu + i \int_{\Sigma} g d^2z X^\mu J_\mu \right\}. \quad (4.2)$$

We decompose the scalars in the zero-mode and the non-zero-mode parts

$$X^\mu = \frac{x^\mu}{\sqrt{\text{Vol}(\Sigma)}} + Y^\mu, \quad \text{with} \quad Y^\mu = \mathbf{P} X^\mu \quad (4.3)$$

where $\mathbf{P}$ is the orthogonal projector on the non-zero-modes

$$\mathbf{P}\psi(z, \bar{z}) = \int g(w, \bar{w}) d^2w P(z, \bar{z}; w, \bar{w}) \psi(w, \bar{w}) \quad (4.4)$$

with integral kernel

$$P(z, \bar{z}; w, \bar{w}) = g(w, \bar{w})^{-1} \delta^{(2)}(z - w) - \frac{1}{\text{Vol}(\Sigma)}, \quad (4.5)$$

and analogously, we decompose the source as

$$J_\mu = \frac{p_\mu}{\sqrt{\text{Vol}(\Sigma)}} + \mathbf{P} J_\mu. \quad (4.6)$$

Then

$$\begin{aligned} Z[J] &= \int dx^\mu e^{ip^\mu x_\mu} \times \\ &\times \int [dY^\mu] \exp \left\{ -\frac{1}{2\pi\alpha'} \int_{\Sigma} d^2z \partial Y^\mu \bar{\partial} Y_\mu + i \int_{\Sigma} g d^2z Y^\mu \mathbf{P} J_\mu \right\}. \end{aligned} \quad (4.7)$$

The first integral yields the spacetime momentum conservation delta-function

$$i(2\pi)^d \delta^d(p_\mu) \quad (4.8)$$

where the overall i arises from the Wick rotation of the time component of p_μ as in field theory. The second factor is a Gaussian path integral without zero-modes which yields

$$[\mathbf{det}'(-\partial\bar{\partial})]^{-d/2} \exp\left\{-\frac{1}{2} \int g(z) d^2z g(w) d^2w G(z; w) \mathbf{PJ}(z)^\mu \mathbf{PJ}(w)_\mu\right\} \quad (4.9)$$

where $\mathbf{det}'$ stands for the functional determinant with the zero-modes omitted, and

$$G(z; w) = G(w; z) \quad (4.10)$$

is the (properly normalized) scalar Green's function on Σ which satisfies the PDE

$$-\frac{1}{\pi\alpha'} \partial_z \bar{\partial}_{\bar{z}} G(z; w) = g(z, \bar{z}) P(z, \bar{z}; w, \bar{w}) \equiv \delta^{(2)}(z - w) - \frac{g(z, \bar{z})}{\text{Vol}(\Sigma)}. \quad (4.11)$$

This PDE is not conformal invariant only because the number $\text{Vol}(\Sigma)$ depends on the particular metric. However from the PDE (together with the symmetry (4.10)), we learn that the difference of the Green functions computed with two conformally equivalent metrics $g dz d\bar{z}$ and $g' dz d\bar{z}$ has the form

$$\Delta G(z; w) \equiv G(z; w) - G'(z; w) = f(z, \bar{z}) + f(w, \bar{w}), \quad (4.12)$$

which implies

$$\int g(z) d^2z g(w) d^2w \Delta G(z; w) \mathbf{PJ}^\mu(z) \mathbf{PJ}_\mu(w) = 0, \quad (4.13)$$

so that we may use Green's function computed in *any* convenient conformally equivalent metric. The additional terms $f(z, \bar{z}) + f(w, \bar{w})$ in Green's function drop out because of the overall spacetime momentum conservation whose effect is to set $J_\mu \equiv \mathbf{PJ}_\mu$; cf. Eq. (4.6). By $G(z; w)$ we always mean Green's function computed with a chosen reference metric in the appropriate conformal class.

A typical quantity we wish to compute is the correlation of tachyon vertices

$$\left\langle : e^{ik_1 \cdot X(z_1)} : \quad : e^{ik_2 \cdot X(z_2)} : \quad \dots \quad : e^{ik_s \cdot X(z_s)} : \right\rangle_\Sigma \quad (4.14)$$

which corresponds to a source of the form

$$\mathbf{PJ}^\mu(z) = \sum_{i=1}^s g(z_i)^{-1} k_i^\mu \delta^{(2)}(z - z_i), \quad \sum_i k_i^\mu = 0, \quad (4.15)$$

where the last condition arises from the overall momentum delta-function. Then, formally, the amplitude is

$$Z(\Sigma) \cdot \exp \left(-\frac{1}{2} \sum_{i,j=1}^s G(z_i, z_j) k_i \cdot k_j \right), \quad (4.16)$$

where $Z(\Sigma) \equiv \langle 1 \rangle_\Sigma$ is the scalars' partition function on the surface Σ . However the exponential factor in Eq. (4.16) requires a renormalization of the divergent diagonal term $i = j$ which implements in the path integral context the normal order prescription of the operator formalism (cf. Sect. 2.3.8). In the diagonal terms, one replaces the Green function $G(z, w)$ by its regularization obtained by subtracting its divergence proportional to $\log d(z, w)$, where $d(z, w)$ is the distance between z and w

$$\begin{aligned} G(z, z)_{\text{reg}} &= \lim_{w \rightarrow z} \left(G(z, w) + \alpha' \log d(z, w) \right) = \\ &= \lim_{w \rightarrow z} \left(G(z, w) + \alpha' \log |z - w| \right) + \alpha' \log g(z)^{1/2} = \\ &\stackrel{\text{def}}{=} -2 \log \lambda(z) + \alpha' \log g(z)^{1/2} \end{aligned} \quad (4.17)$$

since for $z \sim w$, $d(z, w)^2 \approx g(z)|z - w|^2$. The function $\lambda(z)$ is defined by Eq. (4.17). With this prescription, the amplitude becomes

$$\begin{aligned} \text{Equation 4.14} &= i (2\pi)^d \delta^d \left(\sqrt{\text{Vol}(\Sigma)} \sum_i k_i \right) \left[\mathbf{det}'(-\partial\bar{\partial}) \right]_\Sigma^{-d/2} \prod_i \lambda(z_i)^{k_i^2} \times \\ &\times \exp \left(-\frac{\alpha'}{4} \sum_i k_i^2 \log g(z_i) \right) \exp \left(-\sum_{i < j} G(z_i, z_j) k_i \cdot k_j \right). \end{aligned} \quad (4.18)$$

The first exponential in the second line just reflects the fact that the conformal weights of the conformal primary : $e^{ik \cdot X}$: are

$$(h, \tilde{h}) = \left(\frac{\alpha' k^2}{4}, \frac{\alpha' k^2}{4} \right) \quad (4.19)$$

as we know from Sect. 2.4. To simplify the expressions, we often use the short-hand

$$C_\Sigma \stackrel{\text{def}}{=} \text{Vol}(\Sigma) \left[\mathbf{det}'(-\partial\bar{\partial}) \right]_\Sigma. \quad (4.20)$$

The determinant of the Laplacian operator, $\left[\mathbf{det}'(-\partial\bar{\partial}) \right]_\Sigma$, is explicitly computed for Σ an arbitrary compact Riemann surface in [1]. For simplicity here we limit ourselves to genus 0 and 1. Another important correlator is

$$\langle \partial X^\mu(z) \partial X^\nu(w) \rangle_\Sigma = \delta^{\mu\nu} \langle 1 \rangle_\Sigma \partial_z \partial_w G(z; w). \quad (4.21)$$

4.1.1 Scalar Amplitudes on World-Sheets with $\chi \geq 0$

We specialize the above expressions to the compact surfaces Σ with $\chi(\Sigma) \geq 0$.

Scalar Correlations on the Sphere

The punctured sphere is conformal to the plane; by the argument around Eqs. (4.12) and (4.13), we may use directly the Green function in $\mathbb{C}$

$$G(z; w) = -\frac{\alpha'}{2} \log |z - w|^2, \quad (4.22)$$

so that

$$\lambda(z) = g(z) = 1, \quad (4.23)$$

and

$$\begin{aligned} & \left\langle : e^{ik_1 \cdot X(z_1)} : \dots : e^{ik_s \cdot X(z_s)} : \right\rangle_{S^2} = \\ & = (2\pi)^d \delta^d \left(\sum_i k_i \right) C_{S^2}^{-d/2} \exp \left(-\frac{\alpha'}{2} \sum_i k_i^2 \log g(z_i)^{1/2} \right) \prod_{1 \leq i < j \leq s} |z_i - z_j|^{\alpha' k_i \cdot k_j}. \end{aligned} \quad (4.24)$$

Scalar Correlations on the Disk

The disk D is biholomorphic to the upper half-plane $\mathcal{H}$ via the Cayley map (2.304).

The *doubling trick* (cf. Sect. 2.6) allows us to replace the upper half-plane with the full plane $\mathbb{C}$ at the price of adding image insertions in the lower half-plane for all operators in the bulk. Hence the half-plane amplitude on the half-space $\mathcal{H}$ in the presence of the source (4.15) corresponds to the sphere amplitude with the source

$$J^\mu(z) = \sum_{i=1}^s k_i^\mu \left(g(z_i)^{-1} \delta^{(2)}(z - z_i) + g(\bar{z}_i)^{-1} \delta^{(2)}(z - \bar{z}_i) \right). \quad (4.25)$$

The Green function is then

$$G(z; w) = -\frac{\alpha'}{2} \log |z - w|^2 - \frac{\alpha'}{2} \log |z - \bar{w}|^2. \quad (4.26)$$

In view of the Schwarz reflection theorem [2], the restriction of $G(z; w)$ to the upper half-plane is the scalar Green function (i.e. the solution to the PDE (4.11)) satisfying the *Neumann* boundary condition along the real axis $\mathbb{R} \equiv \partial\mathcal{H}$

$$\text{Im} \left(\partial_z G(z; w) \Big|_{z=x \in \mathbb{R}} \right) = 0. \quad (4.27)$$

For insertions in the bulk of the upper half-plane, we have

$$\begin{aligned}
\left\langle \prod_{i=1}^n : e^{ik_i(z_i)} : \right\rangle_{\mathcal{H}} &= i (2\pi)^d (2\pi)^d \delta^d(\Sigma_i k_i) C_{\mathcal{H}}^{-d/2} \times \\
&\times \prod_{i=1}^n |z_i - \bar{z}_i|^{\alpha' k_i^2/2} \prod_{i < j} |z_i - z_j|^{\alpha' k_i \cdot k_j} |z_i - \bar{z}_j|^{\alpha' k_i \cdot k_j}.
\end{aligned} \tag{4.28}$$

Note that now

$$\left\langle \partial X^\mu(z) \bar{\partial} X^\nu(\bar{w}) \right\rangle_{\mathcal{H}} = -\eta^{\mu\nu} \frac{\alpha'}{2} \langle 1 \rangle_{\mathcal{H}} \frac{1}{(z - \bar{w})^2} \neq 0 \tag{4.29}$$

which reflects the fact that the boundary condition now mixes the left- and right-movers, so that we have just one copy of the chiral operator algebra; cf. Sect. 2.6.

Operator Insertions on the Boundary In the case of the upper half-plane $\mathcal{H}$, we are mainly interested in the operator inserted on the boundary ($\equiv$ the real axis $\mathbb{R}$) since these are the ones which correspond to states of the *open* string under the operator-state isomorphism; cf. Sect. 2.6. The bulk expression diverges in the limit $\text{Im } z \rightarrow 0$ because, when restricting (4.26) to the real axis, the two terms become equal

$$G(x, y) = -\alpha' \log |x - y|^2, \quad x, y \in \mathbb{R}, \tag{4.30}$$

and the divergence as $x \rightarrow y$ is now twice as big. Therefore we need to redefine the normal order prescription by introducing the *boundary normal order*

$$\star X^\mu(x) X^\nu(y) \star \stackrel{\text{def}}{=} X^\mu(x) X^\nu(y) + 2\alpha' \eta^{\mu\nu} \log |x - y|. \tag{4.31}$$

Then for the insertion of exponentials all on the boundary

$$\left\langle \prod_{i=1}^n \star e^{ik_i X(x_i)} \star \right\rangle_{\mathcal{H}} = i (2\pi)^d \delta^d(\Sigma_i k_i) C_{\mathcal{H}}^{-d/2} \prod_{i < j} |x_i - x_j|^{2\alpha' k_i \cdot k_j}. \tag{4.32}$$

Scalar Correlations on the Real Projective Plane $\mathbb{RP}^2$

The real projective plane $\mathbb{RP}^2$ is the quotient of the sphere $\mathbb{P}^1$ by the antipodal map $\Omega: z \leftrightarrow -1/\bar{z}$ which has no fixed points. Replacing the sphere by the plane as in Eq. (4.22), the method of images yields Green's function

$$G(z; w) = -\frac{\alpha'}{2} \log |z - w|^2 - \frac{\alpha'}{2} \log |1 + z\bar{w}|^2, \tag{4.33}$$

and then the correlations of exponentials are

$$\begin{aligned}
\left\langle \prod_{i=1}^n : e^{ik_i(z_i)} : \right\rangle_{\mathbb{RP}^2} &= i (2\pi)^d (2\pi)^d \delta^d(\Sigma_i k_i) C_{\mathbb{RP}^2}^{-d/2} \times \\
&\times \prod_{i=1}^n |1 + z_i \bar{z}_i|^{\alpha' k_i^2/2} \prod_{i < j} |z_i - z_j|^{\alpha' k_i \cdot k_j} |1 + z_i \bar{z}_j|^{\alpha' k_i \cdot k_j}. \quad (4.34)
\end{aligned}$$

Note 4.1 $\pi_1(\mathbb{RP}^2) \equiv \pi_1(\mathbb{P}^1/\mathbb{Z}_2) = \mathbb{Z}_2$ and the scalar field may return to itself or to *minus* itself when going through a non-trivial loop. The formulae above correspond to the first case. The second situation has a relative minus sign.

Scalar Correlations on the Torus

We see the torus as the plane $\mathbb{C}$ with the double periodic identification

$$z \sim z + 2\pi m + 2\pi \tau n, \quad m, n \in \mathbb{Z}, \quad \tau \equiv \tau_1 + i\tau_2 \in \mathcal{H}. \quad (4.35)$$

The Green function $G(z; w) \equiv G(z - w)$ satisfies the PDE¹

$$-\frac{1}{\pi\alpha'} \partial \bar{\partial} G(z - w) = \delta^{(2)}(z - w) - \frac{1}{8\pi^2 \tau_2}. \quad (4.36)$$

Formally, the Green function $G(z) \equiv G(z; 0)$ may be obtained by the method of images. In order to do that, we first add a term to $G(z)$ so that the new function $\tilde{G}(z)$ satisfies the PDE (4.36) without the constant term in the RHS

$$\tilde{G}(z) \equiv G(z) + \frac{\alpha' (z - \bar{z})^2}{16\pi \tau_2} \quad (4.37)$$

then

$$\tilde{G}(z) \xrightarrow{\text{formally}} -\frac{\alpha'}{2} \sum_{m,n \in \mathbb{Z}} \log |z - 2\pi m + 2\pi \tau n|^2 \quad (4.38)$$

except that the series in the RHS does not converge, and one must regulate it. From the argument around Eqs. (4.12), (4.13), we see that we are free to subtract from the sum any quantity which depends only on $z, \bar{z}$ or $w, \bar{w}$. The subtraction should preserve the periodicity of $G(z; w)$ under (4.35). In **BOX 4.1**, we go through the procedure and get

$$G(z; w) = -\frac{\alpha'}{2} \log \left| \vartheta_1 \left(\frac{z - w}{2}; \tau \right) \right|^2 - \frac{\alpha' (z - w - \bar{z} + \bar{w})^2}{16\pi \tau_2} \quad (4.39)$$

where the theta-function $\vartheta_1(x; \tau)$ is defined in **BOX 4.1**.

¹ Here we use the volume form $idz \wedge d\bar{z} = 2dx \wedge dy$ for $z = x + iy$.

Next we have to compute the normal order function $\lambda(z)$, see Eq. (4.17):

$$\begin{aligned} \log \lambda(z) &= -\frac{1}{2} \lim_{z \rightarrow w} \left(G(z-w) + \alpha' \log |z-w| \right) = \\ &= \frac{\alpha'}{2} \log \left| \frac{\vartheta_1'(0; \tau)}{2} \right| = \frac{3\alpha'}{2} \log |\eta(\tau)| \end{aligned} \quad (4.40)$$

where we used the identity (♣) in **BOX 4.1**. Then (here and below we set $z_{ij} \equiv z_i - z_j$)

$$\begin{aligned} \left\langle \prod_{i=1}^n : e^{ik_i(z_i)} : \right\rangle_{T^2} &= i (2\pi)^d (2\pi)^d \delta^d(\Sigma_i k_i) C(\tau) T_2^{-d/2} \times \\ &\times \prod_{i=1}^n |\eta(\tau)|^{3\alpha' k_i^2/2} \prod_{i < j} \left| \vartheta_1 \left(\frac{z_{ij}}{2}; \tau \right) \right|^{\alpha' k_i \cdot k_j} \exp \left[-\frac{\text{Im}(z_{ij})}{4\pi \tau_2} \alpha' k_i \cdot k_j \right]. \end{aligned} \quad (4.41)$$

The Scalar Partition Function on the Torus

For the torus the quantity

$$C(\tau)_{T^2} \equiv \text{Vol}(T^2) \mathbf{det}'[-\partial \bar{\partial}]_{T^2} \quad (4.42)$$

is no longer a constant but a function of the conformal modulus $\tau \equiv \tau_1 + i\tau_2$ of T^2 . The volume factor is $8\pi^2 \tau_2$. The second factor is related to the partition function of d free real scalars by

$$Z(\tau) = i \mathbf{det}'[-\partial \bar{\partial}]_{T^2}^{-d/2}. \quad (4.43)$$

The torus of period $\tau_1 + i\tau_2$ is obtained by gluing together the two circular boundaries of a cylinder of circumference 2π and length $2\pi\tau_2$ with the identification $\theta' \sim \theta + 2\pi\tau_1$. The partition function on T^2 is then the trace over the Hilbert space of the Euclidean time evolution $\exp[-2\pi\tau_2 H]$ times the translation $\exp[2\pi i\tau_1 P]$

$$Z(\tau) = \text{Tr}_{\mathbf{H}} \left[e^{-2\pi\tau_2 H} e^{2\pi i\tau_1 P} \right], \quad (4.44)$$

with $\mathbf{H}$ the Hilbert space of d scalars with periodic b.c. $X^\mu(z + 2\pi) = X^\mu(z)$.

In a CFT the cylinder Hamiltonian H and momentum operator P are

$$H = L_0 + \tilde{L}_0 - \frac{c + \tilde{c}}{24}, \quad P = L_0 - \tilde{L}_0, \quad (4.45)$$

so we get

$$Z(\tau) = \text{Tr} \left[q^{L_0 - d/24} \bar{q}^{\tilde{L}_0 - d/24} \right], \quad q \equiv e^{2\pi i \tau} \quad (4.46)$$

BOX 4.1 - The scalar Green's functions on the torus

We recall the identities [3]

$$\begin{aligned} \sum_{m \in \mathbb{Z}} \frac{\partial}{\partial x} \log(x - \pi m) &= \frac{\partial}{\partial x} \log x + \sum_{m \geq 1} \frac{\partial}{\partial x} \log(x^2 - \pi^2 m^2) = \\ &= \frac{1}{x} + \sum_{m \geq 1} \frac{2x}{x^2 - \pi^2 m^2} = \frac{\cos x}{\sin x} = \frac{\partial}{\partial x} \log \sin x = i + \frac{\partial}{\partial x} \log(1 - e^{-2ix}). \end{aligned}$$

Hence ($q \equiv e^{2\pi i \tau}$ as always)

$$\begin{aligned} \sum_{m, n \in \mathbb{Z}} \frac{\partial^2}{\partial z^2} \log(z - 2\pi m + 2\pi \tau n) &= \sum_{n \in \mathbb{Z}} \frac{\partial^2}{\partial z^2} \log(1 - e^{-2\pi i \tau n} e^{-iz}) = \\ &= \frac{\partial^2}{\partial z^2} \log \sin \frac{z}{2} + \sum_{n \geq 1} \frac{\partial^2}{\partial z^2} \left[\log(1 - q^n e^{iz}) + \log(1 - q^n e^{-iz}) \right] = \frac{\partial^2}{\partial z^2} \log \vartheta_1\left(\frac{z}{2}; \tau\right) \end{aligned}$$

where (in our conventions) the ϑ -function is given by the Jacobi triple product as

$$\vartheta_1(x; \tau) = 2q^{1/8} \sin x \prod_{n=1}^{\infty} (1 - e^{2ix} q^n)(1 - e^{-2ix} q^n). \quad (\text{Jacobi})$$

Therefore

$$\frac{\partial^2}{\partial z^2} \tilde{G}(z) = -\frac{\alpha'}{2} \frac{\partial^2}{\partial z^2} \log \vartheta_1\left(\frac{z}{2}; \tau\right) \Rightarrow \tilde{G}(z) = -\frac{\alpha'}{2} \log \left| \vartheta_1\left(\frac{z}{2}; \tau\right) \right|^2 + az + \bar{a}\bar{z} + c$$

for some integration constants a and c . The amplitudes do not depend on a, c by momentum conservation, and we may set $a = c = 0$. Then

$$G(z) = -\frac{\alpha'}{2} \log \left| \vartheta_1\left(\frac{z}{2}; \tau\right) \right|^2 - \frac{\alpha' (z - \bar{z})^2}{16\pi \tau_2} \quad (\star)$$

$\vartheta_1(z; \tau)$ has the (pseudo)periodicity property

$$\left| \vartheta_1(x + (m + n\tau)\pi; \tau) \right|^2 = e^{\pi n^2(i\bar{\tau} - i\tau)} e^{2n(i\bar{x} - ix)} \left| \vartheta_1(x; \tau) \right|^2$$

which implies that $(\star)$ is doubly periodic, as it should be. From Eq. (Jacobi) we see that

$$\left. \partial_x \vartheta_1(x; \tau) \right|_{x=0} = 2q^{1/8} \prod_{n=1}^{\infty} (1 - q^n)^3 = 2\eta(\tau)^3 \quad (\clubsuit)$$

where $\eta(\tau)$ is the Dedekind function $\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n)$, a holomorphic function without zeros in $\mathcal{H}$ with specific modular properties (see Eq. (4.58)).

where (cf. Sect. 2.4)

$$L_0 = \frac{\alpha'}{4} p^\mu p_\mu + N \quad N = \sum_{n \geq 1} \alpha_{-n}^\mu \alpha_{n\mu} \quad (4.47)$$

$$\tilde{L}_0 = \frac{\alpha'}{4} p^\mu p_\mu + \tilde{N} \quad \tilde{N} = \sum_{n \geq 1} \tilde{\alpha}_{-n}^\mu \tilde{\alpha}_{n\mu}, \quad (4.48)$$

and $N, \tilde{N}$ are the left/right oscillator numbers

$$N = \sum_{\mu=0}^{d-1} \sum_{n \geq 1} n N_{\mu n}, \quad \tilde{N} = \sum_{\mu=0}^{d-1} \sum_{n \geq 1} n \tilde{N}_{\mu n}, \quad (4.49)$$

where $N_{\mu n}$ (resp. $\tilde{N}_{\mu n}$) is the occupation number for the oscillator α_{-n}^{μ} (resp. $\tilde{\alpha}_{-n}^{\mu}$). The trace breaks into a sum over the occupation numbers $N_{\mu n}, \tilde{N}_{\mu n}$ and an integral over the momentum k_{μ} . The continuum normalization of the momentum yields

$$\sum_k \rightsquigarrow V_d \int \frac{d^d k}{(2\pi)^d}, \quad (4.50)$$

where V_d is the volume of spacetime. Then

$$Z(\tau) = (q\bar{q})^{-d/24} V_d \int \frac{d^d k}{(2\pi)^d} \exp(-\pi \tau_2 \alpha' k^2) \prod_{\mu, n} \sum_{\substack{N_{\mu n}=0 \\ \tilde{N}_{\mu n}=0}}^{\infty} q^{n N_{\mu n}} \bar{q}^{n \tilde{N}_{\mu n}}. \quad (4.51)$$

The integral is Gaussian and yields a factor

$$i (4\pi^2 \tau_2 \alpha')^{-d/2}, \quad (4.52)$$

while the sums are geometric series

$$\sum_{N=0}^{\infty} q^{nN} = (1 - q^n)^{-1} \quad (4.53)$$

so that

$$Z(\tau) = i V_d Z_X(\tau)^d \quad (4.54)$$

where $Z_X(\tau)$ is the *partition function of one real non-compact scalar*:

$$Z_X(\tau) \stackrel{\text{def}}{=} (4\pi^2 \alpha' \tau_2)^{-1/2} |\eta(\tau)|^{-2}, \quad (4.55)$$

and $\eta(\tau)$ is the *Dedekind function*

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n), \quad q \equiv e^{2\pi i \tau}. \quad (4.56)$$

The infinite product in the RHS converges for $|q| < 1$, so $\eta(\tau)$ is a holomorphic function in the upper half-plane $\mathcal{H}$, without zeros for finite τ , which enjoys nice

modular properties: under $\tau \rightarrow \tau + 1$, $\eta(\tau)$ changes by a phase (a 24th-root of unity)² which cancels in $Z_X(\tau)$. One has the identities [3, 5, 6]

$$\begin{aligned}\eta(\tau + 1) &= e^{2\pi i/24} \eta(\tau), \\ \eta(-1/\tau) &= (-i\tau)^{1/2} \eta(\tau), \\ \text{Im}(-1/\tau) &= \frac{\tau_2}{\tau\bar{\tau}},\end{aligned}\tag{4.57}$$

so that

$$Z_X(\tau + 1) = Z_X(\tau), \quad Z_X(-1/\tau) = Z_X(\tau),\tag{4.58}$$

as expected since the partition function of a well-defined CFT is modular invariant; cf. Sect. 2.3.7. The transformations $T: \tau \mapsto \tau + 1$ and $S: \tau \mapsto -1/\tau$ generate the full modular group $PSL(2, \mathbb{Z})$ so that

$$Z_X\left(\frac{a\tau + b}{c\tau + d}\right) = Z_X(\tau) \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}).\tag{4.59}$$

We also recall the two identities

$$\eta(\tau) = q^{1/24} \sum_{n=-\infty}^{+\infty} (-1)^n q^{(3n^2+n)/2}, \quad \eta(\tau)^{-1} = q^{-1/24} \sum_{n \geq 0} P(n) q^n\tag{4.60}$$

where $P(n)$ is the number of partitions of n ; cf. Sect. 2.3.4.

Scalar Green's Function on the Annulus with Neumann b.c.

We see the annulus as the finite flat cylinder Cy_t obtained by identifying the strip $0 \leq \text{Re } z \leq \pi$ periodically in “time”, $z \sim z + 2\pi it$, where $t \in \mathbb{R}$ is the *real* modulus of the finite cylinder.³ The doubling trick identifies Cy_t with the quotient of its closed double E_{it} by an orientation-reversing involution with fixed set ∂Cy_t . The closed double E_{it} is the rectangular torus E_{it} with modulus $\tau = it$,

$$E_{it} \equiv \mathbb{C} / (z \sim z + 2\pi m + 2\pi i t n), \quad m, n \in \mathbb{Z}\tag{4.61}$$

$$\text{Cy}_t \equiv E_{it} / (z \sim 2\pi - \bar{z});\tag{4.62}$$

see Fig. 4.1a, b. The image method reduces us to the torus E_{it} with a set of insertions invariant under the reflection $z \leftrightarrow 2\pi - \bar{z}$. As in the disk, the Green function with *Neumann boundary condition* is obtained by inserting a mirror image of the source

$$G(z; w)_{\text{cylinder}} = G(z; w)_{\text{torus } \tau=it} + G(z; -\bar{w})_{\text{torus } \tau=it}.\tag{4.63}$$

² Hence the function $\Delta(\tau) \equiv (2\pi)^{12} \eta(\tau)^{24}$ —called the *discriminant* [4]—is a good weight 12 *cusp form* (i.e. a weight 12 holomorphic modular form which vanishes at infinity).

³ By the Riemann–Roch theorem and $\chi = 0$, we know that the number of real moduli is equal to the number of real CKV. There is *one* CKV generating translations in periodic Euclidean “time”.

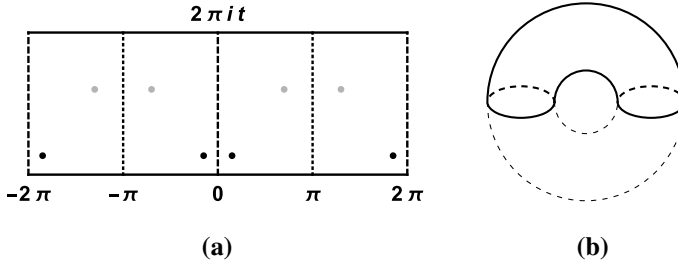


Fig. 4.1 **a** The annulus as a quotient of the torus by an anti-holomorphic involution. Solid horizontal lines and vertical dashed (resp. dotted) lines are periodically identified. Equivalent points have equal color. Vertical dashed (resp. dotted) lines (which are fixed by the involution) are the two boundaries. **b** An alternative view of the finite cylinder ($\equiv$ annulus) as a quotient of the rectangular torus (cf. Fig. 4.1a). The torus is portrayed as “a donut” in $\mathbb{R}^3$. The involution (4.62) is the reflection in a horizontal plane passing through the center of the figure

As with the disk, we have to distinguish between operators inserted in the bulk of the cylinder and operators inserted on the boundaries $\text{Re } z = 0, \pi$. The normal order prescription in the two cases differs for the same reason as in the disk. We leave the details as an **Exercise**.

The Partition Function on the Finite Cylinder

The finite cylinder ($\equiv$ annulus) path integral computes the trace of $\exp(-2\pi t H)$ on the Hilbert space of the free scalar theory quantized on the segment $[0, \pi]$ with Neumann boundary conditions (open sector of the string)

$$\text{Tr}_{\text{open}} \left[e^{-2\pi t (L_0 - d/24)} \right]. \quad (4.64)$$

In the open sector $L_0 = \alpha' p^2 + N$, and we get

$$Z(t)_{\text{cylinder}} = i V_d (8\pi^2 \alpha' t)^{-d/2} \eta(it)^{-d}. \quad (4.65)$$

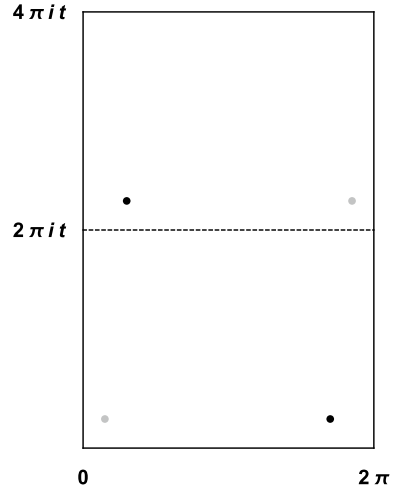
This path integral has an alternative interpretation. Indeed, we may see the circle in the cylinder $[0, \pi] \times S^1$ as “space” and the segment $[0, \pi]$ as “Euclidean time”: from this viewpoint the finite cylinder represents the *tree-level* propagation of a *closed* string between two states defined by the boundary conditions on the circles at the two ends (the Neumann b.c. in our case): such states are called *boundary states*; see Sect. 4.8 below. Thus, schematically,

$$Z(t)_{\text{cylinder}} = \langle \text{Neumann} | e^{-\pi H} | \text{Neumann} \rangle. \quad (4.66)$$

Scalar Green’s Function in the Klein Bottle

The Klein bottle KI is described in **BOX 1.5**. We may equivalently describe KI as the quotient of its oriented double—the rectangular torus E_{2it} of periods $(2\pi, 4\pi it)$ ($t \in \mathbb{R}$)—by the orientation-reversing involution without fixed points

Fig. 4.2 Fundamental domain of Ω acting on the torus of periods $(2\pi, 4\pi it)$ is the rectangle R in Eq. (4.68), with the opposite vertical sides $x = 0$ and $x = 2\pi$ identified, making a cylinder whose two S^1 boundaries, $y = 0$ and $y = 2\pi t$, are identified according to the rule $(x, 0) \sim (-x, 2\pi t)$, that is, with a reversal of orientation. Same color points are identified



$$\Omega: z \sim -\bar{z} + 2\pi it. \quad (4.67)$$

The fundamental domain of the action of Ω is the rectangle

$$R = \{0 \leq x \leq 2\pi, 0 \leq y \leq 2\pi t\} \quad (4.68)$$

with the opposite vertical sides $x = 0$ and $x = 2\pi$ identified, making a cylinder whose two S^1 boundary components, $y = 0$ and $y = 2\pi t$, are identified with the rule

$$(x, 0) \sim (-x, 2\pi t), \quad (4.69)$$

that is, with a *reversal of orientation*; see **Fig. 4.2**. In other words, both the torus and the Klein bottle are obtained by gluing the two circular ends of a cylinder, but in the second case we twist the orientation of a boundary before gluing. We write Ω for the unitary operator acting on the Hilbert space in S^1 which inverts the orientation.

Again Green's function is obtained by the method of images

$$G(z; w)_{\text{Klein}} = G(z; w)_{\text{torus } \tau=2it} + G(z; -\bar{w} + 2\pi it)_{\text{torus } \tau=2it}. \quad (4.70)$$

The Partition Function on the Klein Bottle

The above description of the Klein bottle as a cylinder with its ends identified with the orientation twist Ω yields the path integral as a trace over the closed sector Hilbert space $\mathbf{H}$

$$Z(\tau)_{\text{Klein}} = \text{Tr}_{\mathbf{H}} \left[e^{-2\pi t(L_0 + \bar{L}_0 - d/12)} \Omega \right]. \quad (4.71)$$

The scalars' zero-modes are left-right symmetric and contribute a factor

$$i V_d (4\pi^2 \alpha' t)^{-d/2} \quad (4.72)$$

as in the torus case. Because of the insertion of Ω , only states which are left-right symmetric contribute to the trace, and hence the sum over oscillator numbers in Eq. (4.51) gets restricted to $N_{\mu n} = \tilde{N}_{\mu n}$. Then the oscillator modes produce the factor

$$e^{-4\pi t d/24} \prod_{\mu, n} \sum_{N_{\mu n}=0}^{\infty} e^{-4\pi i t n N_{\mu n}} = \eta(2it)^{-d}, \quad (4.73)$$

and finally

$$Z(t)_{\text{Klein}} = i V_d (4\pi^2 \alpha' t)^{-d/2} \eta(2it)^{-d}. \quad (4.74)$$

Partition Function for the Möbius Strip with Neumann b.c.

We visualize the Möbius strip Mö as the rectangle

$$R = \{0 \leq x \leq \pi, 0 \leq y \leq 2\pi t\} \quad (4.75)$$

where we identify the two horizontal sides with an inversion Ω of the orientation

$$\Omega: (x, 0) \sim (\pi - x, 2\pi t). \quad (4.76)$$

The partition function is then a trace over the open-sector Hilbert space

$$Z(t)_{\text{Möbius}} = \text{Tr}_{\text{open}} \left[e^{-2\pi t (L_0 - d/24)} \Omega \right]. \quad (4.77)$$

The unitary operator Ω satisfies (cf. (3.305))

$$\Omega \alpha_n^\mu \Omega^{-1} = (-1)^n \alpha_n^\mu, \quad n \in \mathbb{Z}. \quad (4.78)$$

Hence the contribution to the trace in (4.77) from non-zero-modes is

$$e^{2\pi t d/24} \prod_{\mu, n \geq 1} \sum_{N_{\mu n}=0}^{\infty} (-1)^{n N_{\mu n}} e^{-2\pi t n N_{\mu n}} = \left(e^{-2\pi t/24} \prod_{n=1}^{\infty} (1 - (-e^{-2\pi t})^n) \right)^{-d}. \quad (4.79)$$

One has

$$e^{-2\pi t/24} \prod_{n=1}^{\infty} (1 - (-e^{-2\pi t})^n) = e^{2\pi t/24} \prod_{n=1}^{\infty} (1 - e^{-4\pi t n})(1 + e^{-4\pi t n + 2\pi t}). \quad (4.80)$$

We recall the Jacobi triple product ($q \equiv e^{2\pi i \tau}$) [3, 7]

$$\vartheta_3(z; \tau) = \prod_{n \geq 1} (1 - q^n)(1 + e^{iz} q^{n-1/2})(1 + e^{-iz} q^{n-1/2}) \quad (4.81)$$

which allows to write

$$\begin{aligned} e^{-2\pi t/24} \prod_{n=1}^{\infty} (1 - (-e^{-2\pi t})^n) &= e^{-2\pi t/24} \prod_{n=1}^{\infty} (1 - e^{-4\pi t n})^{1/2} \vartheta_3(0, 2it)^{1/2} = \\ &= \eta(2it)^{1/2} \vartheta_3(0; 2it)^{1/2}. \end{aligned} \quad (4.82)$$

Then

$$Z(t)_{\text{Möbius}} = i V_d (8\pi^2 \alpha' t)^{-d/2} \vartheta_3(0, 2it)^{-d/2} \eta(2it)^{-d/2}. \quad (4.83)$$

4.2 Amplitudes for the b, c CFT

The ghost system is Gaussian, so the path integral is a functional determinant. The only subtle point is the soaking of zero-modes: as discussed in Chap. 1, their net effect is to produce the correct measure on the complex moduli space $\mathcal{M}_{g,n}$. For the applications to string theory, we need only the amplitudes with the ghost insertions required to cancel the zero-modes since the only purpose of the b, c ghosts is to produce the right functional measure in the covariant quantization of the string.

Note 4.2 The path integrals for free b, c chiral systems of *any* conformal spin λ (cf. Sect. 2.5) on Riemann surfaces of *any* genus g can be explicitly computed using bosonization techniques together with deep geometric facts; see [1, 8–10].

Ghost Amplitudes on the Sphere $S^2 \equiv \mathbb{P}^1$ has 3 complex CKV and is rigid. Hence we have 3 c zero-modes $v_i(z)$, 3 $\tilde{c}$ zero-modes $\tilde{v}_i(\bar{z})$, and no b or $\tilde{b}$ zero-modes. By Fermi statistics, the non-zero amplitude is

$$\langle c(z_1)c(z_2)c(z_3)\tilde{c}(\bar{w}_1)\tilde{c}(\bar{w}_2)\tilde{c}(\bar{w}_3) \rangle_{S^2} = C \det[v_i(z_j)] \det[\tilde{v}_i(\bar{w}_j)] \quad (4.84)$$

where C is an overall constant (equal to⁴ $|\det' \bar{\partial}|^2$) and $v_i(z)$ ($i = 1, 2, 3$) is an orthonormal basis of the algebra $\mathfrak{sl}(2, \mathbb{C})$ of $(1,0)$ CKVs. We may use any basis at the cost of redefining the overall constant C . From Eq. (CKV) in **BOX 1.11**, we see that we may take $v_i(z) = z^{i-1}$. Then

$$\begin{aligned} \langle c(z_1)c(z_2)c(z_3)\tilde{c}(\bar{w}_1)\tilde{c}(\bar{w}_2)\tilde{c}(\bar{w}_3) \rangle_{S^2} &= \\ &= C' \det(z_j)^{i-1} \det[(\bar{w}_j)^{i-1}] = C' \prod_{i < j} (z_i - z_j)(\bar{w}_i - \bar{w}_j) \end{aligned} \quad (4.85)$$

where we used the Vandermonde determinant formula.

⁴ As always a *prime* on a determinant means the product of non-zero eigenvalues.

Comparison with Bosonization In the above formula, we choose $w_i = z_i$ and consider the amplitude

$$\left\langle \prod_{i=1}^3 e^{\Phi(z_i, \bar{z}_i)} \right\rangle_{S^2} \quad (4.86)$$

where $\Phi(z, \bar{z}) = \phi(z) + \tilde{\phi}(\bar{z})$ is the non-chiral scalar obtained by combining the left- and right-moving scalars which bosonize the $b, c, \tilde{b}$, and $\tilde{c}$ systems. Its OPE is

$$\Phi(z) \Phi(w) \sim -\log |z - w|^2 \quad (4.87)$$

and its action has the standard free form plus a linear term (cf. Sect. 2.5.5)

$$-3 \int \Phi \frac{R}{8\pi^2} d^2 z. \quad (4.88)$$

Taking the metric to have constant scalar curvature, this term affects only the zero-mode part of the scalar amplitude, and its net effect is to change the charge (“momentum”) conservation $\delta(\sum_i q_i)$ to $\delta(\sum_i q_i - 3)$. Omitting the delta-function, we have

$$\left\langle \prod_{i=1}^3 e^{\Phi(z_i, \bar{z}_i)} \right\rangle_{S^2} = \text{const.} \prod_{i < j} |z_i - z_j|^2, \quad (4.89)$$

in agreement with the previous result.

Ghost Amplitudes on the Disk The CKV algebra on the upper half-plane $\mathcal{H}$ is $\mathfrak{sl}(2, \mathbb{R})$, and we have 3 real zero-modes of c and no zero-modes of b since $\mathcal{H}$ is rigid. The doubling trick extends $b(z), c(z)$ holomorphically to the full plane $\mathbb{C}$ by setting

$$b(z) = \tilde{b}(\bar{z}), \quad c(z) = \tilde{c}(\bar{z}) \quad \text{for } \text{Im } z < 0. \quad (4.90)$$

The most convenient way to soak up the 3 real zero-modes is to insert the ghost field c at 3 distinct points x_i along the real axis (cf. Sect. 1.6); then

$$\left\langle c(x_1) c(x_2) c(x_3) \right\rangle_{\mathcal{H}} = C'' \prod_{i < j} (x_i - x_j). \quad (4.91)$$

Ghost Amplitudes on the Real Projective Plane $\mathbb{RP}^2$ is rigid, so no b or $\tilde{b}$ zero-modes. $\mathbb{RP}^2$ is the quotient of $\mathbb{P}^1$ under the anti-holomorphic involution

$$z \mapsto w \equiv -1/\bar{z}, \quad (4.92)$$

so the CKV on $\mathbb{RP}^2$ are the CKV on $\mathbb{P}^1$

$$v(z) \partial_z + \tilde{u}(\bar{z}) \partial_{\bar{z}} \equiv (a + bz + cz^2) \partial_z + (d + e\bar{z} + f\bar{z}^2) \partial_{\bar{z}} \quad (4.93)$$

such that

$$v(-1/\bar{w}) \frac{\partial \bar{w}}{\partial \bar{z}} \partial_{\bar{w}} + \tilde{u}(-1/\bar{w}) \frac{\partial w}{\partial \bar{z}} \partial_w = v(w) \partial_w + \tilde{u}(\bar{w}) \partial_{\bar{w}}, \quad (4.94)$$

that is,

$$\tilde{u}(\bar{w}) = \bar{w}^2 v(-1/\bar{w}) \quad (4.95)$$

so that we have the 3 CKV

$$K_{-1} = \partial_z + \bar{z}^2 \partial_{\bar{z}}, \quad K_0 = z \partial_z - \bar{z} \partial_{\bar{z}}, \quad K_{+1} = (z^2 \partial_z + \partial_{\bar{z}}) \quad (4.96)$$

which generate the Lie algebra $\mathfrak{su}(2)$. The ghost fields may be extended to the full $\mathbb{P}^1$ by the doubling trick by imposing the same conditions

$$\tilde{c}(\bar{w}) = \bar{w}^2 c(-1/\bar{w}), \quad \tilde{b}(\bar{w}) = (\bar{w})^{-4} b(-1/\bar{w}). \quad (4.97)$$

Using the image trick, the amplitude reduces to a correlation on the sphere, and

$$\langle c(z_1) c(z_2) c(z_3) \rangle_{\mathbb{R}P^2} = C''' \prod_{i < j} (z_i - z_j)(1/\bar{z}_i - 1/\bar{z}_j). \quad (4.98)$$

Ghost Amplitudes on the Torus The torus has a complex modulus and a complex CKV. Thus we have one zero-mode for each of the fields $b(z)$, $\tilde{b}(\bar{z})$, $c(z)$, and $\tilde{c}(\bar{z})$. The path integral on the torus is a trace over the Hilbert space. For fermions the trace $\text{Tr}(e^{-\beta H})$ (the partition function) is given by the path integral with *anti-periodic* boundary conditions in Euclidean time [11]. The periodic Euclidean path integral yields $\text{Tr}[(-1)^F e^{-\beta H}]$ where $(-1)^F$ is the *fermion parity operator* which counts fermions mod 2 [12].

In our case b, c are Faddeev–Popov ghosts which should have the same periodicity conditions as the gauge constraint to which they refer, here the energy–momentum tensor, so they are periodic along the cycles of the torus T^2 . Therefore the relevant b, c amplitude is

$$\begin{aligned} \text{Tr} \left[(-1)^F c_0 b_0 \tilde{c}_0 \tilde{b}_0 e^{2\pi i \tau_1 P - 2\pi \tau_2 H} \right] &= \\ &= (q\bar{q})^{26/24} \text{Tr} \left[(-1)^F c_0 b_0 \tilde{c}_0 \tilde{b}_0 q^{L_0} \bar{q}^{\bar{L}_0} \right]. \end{aligned} \quad (4.99)$$

Clearly without the zero-mode insertions, we would get zero because of the four-fold degeneracy of the vacuum on the cylinder. The insertion projects out the states obtained by acting with the non-zero oscillator modes on 3 of the 4 vacua.

We recall the form of L_0 for the b, c system

$$\begin{aligned}
L_0 &= \sum_{m \in \mathbb{Z}} m : b_{-m} c_m : = \sum_{m \geq 1} m (b_{-m} c_m + c_{-m} b_m) - 1 = \\
&= \sum_{m \geq 1} (m N_{c,m} + m N_{b,m}) - 1,
\end{aligned} \tag{4.100}$$

where $N_{c,m}$, $N_{b,m}$ are the occupation numbers of the m -th harmonic of the field c , resp. b which for fermions can have only the values 0 or 1. Then

$$\begin{aligned}
\text{Equation 4.99} &= (q\bar{q})^{1/12} \prod_{n \geq 1} \sum_{\substack{N_{c,n}, N_{b,n} \\ N_{\tilde{c},n}, N_{\tilde{b},n}}} (-q^n)^{(N_{c,n} + N_{b,n})} (-\bar{q}^n)^{(N_{\tilde{c},n} + N_{\tilde{b},n})} \\
&= (q\bar{q})^{1/12} \prod_{n \geq 1} (1 - q^n)^2 (1 - \bar{q}^n)^2 = |\eta(\tau)|^4.
\end{aligned} \tag{4.101}$$

Ghost Amplitude on the Cylinder In this case we have one real modulus and one real CKV, so b and c have one zero-mode each ($\tilde{b}$, $\tilde{c}$ are identified with b , c by the doubling trick). The computation is similar to the torus one, except that $\tau = it$ with $t \in \mathbb{R}$ and we have only one copy of the b , c system. Then the b , c partition function (with zero-modes soaked up) on the cylinder is

$$\langle b(z) c(z) \rangle_{\text{cylinder}} = \eta(it)^2. \tag{4.102}$$

Ghost Amplitude on the Klein Bottle Again we have one zero-mode for b and one for c ($\tilde{b}$ and $\tilde{c}$ are identified by going to the oriented double T^2). Arguing as in Eqs. (4.71)–(4.74), we get for the partition function (with zero-modes soaked up) on the Klein bottle

$$\langle b(z) c(z) \rangle_{\text{Klein}} = \eta(2it)^2. \tag{4.103}$$

Ghost Amplitude on the Möbius Strip Again we have one b zero-mode and one c zero-mode and the relevant amplitude is

$$\langle c(z) b(w) \rangle_{\text{Möbius}} = \text{Tr}_{\text{open}} \left[(-1)^F c_0 b_0 e^{-2\pi i t (L_0 + 13/12)} \Omega \right]. \tag{4.104}$$

Exercise 4.1 Show that $\Omega b_n \Omega^{-1} = (-1)^n b_n$, and $\Omega c_n \Omega^{-1} = (-1)^n c_n$.

Then

$$\begin{aligned}
\text{Equation 4.104} &= e^{-2\pi t/12} \prod_{n \geq 1} \sum_{N_{c,n}, N_{b,n}} \left(-(-e^{-2\pi t})^n \right)^{N_{c,n} + N_{b,n}} = \\
&= \eta(2it) \vartheta_3(0; 2it)
\end{aligned} \tag{4.105}$$

where we used the identity (4.82).

Summary of Ghost Amplitudes

In all surfaces Σ the determinants *over non-zero-modes* Δ' are related as

$$\Delta' \Big|_{b,c \text{ system}} = \left(\Delta' \Big|_{\text{real scalar } X} \right)^2, \quad (4.106)$$

i.e. the net effect of the non-zero-modes of b, c is to cancel the oscillators of the two longitudinal scalars $X^\pm$ leaving with $d - 2$ transverse scalars (as expected by equivalence with the light-cone quantization), while their zero-modes reproduce the correct measure on moduli.

4.3 The Veneziano Amplitude

In the rest of the chapter, we compute explicitly a few sample (important) physical amplitudes in the bosonic string. We first consider tree-level amplitudes for the oriented open string.

Three Tachyon Disk Amplitude

We identify the disk with the upper half-plane $\mathcal{H}$. The disk contribution to the three open string tachyon S -matrix element is

$$g_o^3 e^{-\lambda} \left\langle {}^*c e^{ik_1 X}(x_1) {}^* {}^*c e^{ik_2 X}(x_2) {}^* {}^*c e^{ik_3 X}(x_3) {}^* \right\rangle_{\mathcal{H}} + (k_2 \leftrightarrow k_3). \quad (4.107)$$

g_o is the normalization factor of the tachyon vertex (proportional to the open string coupling). The factor $e^{-\lambda}$ originates from the Euler term in the Polyakov action using $\chi(\mathcal{H}) = 1$. The two terms in Eq. (4.107) arise from the sum over the two cyclic orders of the three vertices on the boundary

$$\partial\mathcal{H} = \mathbb{R} \cup i\infty \simeq S^1; \quad (4.108)$$

this sum enforces the spacetime Bose symmetry. Using (4.32) and (4.91), we get

$$i g_o^3 C (2\pi)^{26} \delta^{26}(\Sigma_i k_i) \prod_{i < j} |x_i - x_j|^{1+2\alpha' k_i k_j} + (k_2 \leftrightarrow k_3) \quad (4.109)$$

where the overall constant is

$$C = e^{-\lambda} C'' C_{\mathcal{H}}^{-13}. \quad (4.110)$$

Momentum conservation and the mass-shell condition $k_i^2 = 1/\alpha'$ imply

$$2\alpha' k_1 \cdot k_2 = \alpha' (k_3^2 - k_1^2 - k_2^2) = -1 \quad (4.111)$$

so that the amplitude reduces to

$$2i g_0^3 C (2\pi)^{26} \delta^{26}(\Sigma_i k_i) \quad (4.112)$$

independently of the choice of the x_i 's as a consequence of the Faddeev–Popov procedure. Indeed, we know that by $SL(2, \mathbb{R})$ symmetry we may fix the three operator insertions in three points of our choice along the real line. Weyl invariance is crucial here: the amplitude is independent of the points x_i only on-shell.⁵

The Four Tachyon Amplitude

The 4 open string tachyon amplitude on the disk is

$$\begin{aligned} & g_0^4 e^{-\lambda} \int_{-\infty}^{+\infty} dx_4 \left\langle \star e^{ik_4 \cdot X(x_4)} \star \prod_{i=1}^3 \star c(x_i) e^{ik_i \cdot X(x_i)} \star \right\rangle_H + (k_2 \leftrightarrow k_3) = \\ & = i g_0^4 (2\pi)^{26} \delta^{26}(\Sigma_i k_i) |x_{12} x_{13} x_{23}| \int_{-\infty}^{+\infty} dx_4 \prod_{i < j} |x_{ij}|^{2\alpha' k_i \cdot k_j} + (k_2 \leftrightarrow k_3) \end{aligned} \quad (4.113)$$

where we use the short-hand notation $x_{ij} \equiv x_i - x_j$. Again, the amplitude is independent of x_1 , x_2 and x_3 , which are conveniently fixed at (respectively) 0, 1, and ∞ . We parametrize the amplitude in terms of the *Mandelstam variables*

$$s \equiv -(k_1 + k_2)^2, \quad t \equiv -(k_1 + k_3)^2, \quad u \equiv -(k_1 + k_4)^2 \quad (4.114)$$

which are related by

$$s + t + u = - \sum_i k_i^2 = \sum_i m_i^2 = -\frac{4}{\alpha'}. \quad (4.115)$$

Since

$$2\alpha' k_i \cdot k_j = -2 + \alpha' (k_i + k_j)^2, \quad (4.116)$$

the amplitude takes the form

$$i g_0^4 C (2\pi)^{26} \delta^{26}(\Sigma_i k_i) \left[\int_{-\infty}^{+\infty} dx_4 |x_4|^{-\alpha' u - 2} |1 - x_4|^{-\alpha' t - 2} + (t \leftrightarrow s) \right]. \quad (4.117)$$

The integral is easily computed in terms of the Euler beta-function⁶

⁵ Recall that vertices are primaries of the right dimensions only for on-shell momenta.

⁶ In computing the integral we set $x = 1 - y^{-1}$ and $v = 1/x$.

$$\begin{aligned}
& \int_{-\infty}^{+\infty} dx |x|^{a-1} |1-x|^{b-1} = \\
& = \int_{-\infty}^0 dx |x|^{a-1} |1-x|^{b-1} + \int_0^1 dx |x|^{a-1} |1-x|^{b-1} + \int_1^{\infty} dx |x|^{a-1} |1-x|^{b-1} = \quad (4.118) \\
& = \int_0^1 dy y^{-a-b} (1-y)^{a-1} + \int_0^1 dx x^{a-1} (1-x)^{b-1} + \int_0^1 dv v^{-a-b} (1-v)^{b-1} = \\
& = B(1-a-b, a) + B(a, b) + B(1-a-b, b),
\end{aligned}$$

where

$$\begin{aligned}
B(a, b) & \stackrel{\text{def}}{=} \int_0^1 dx x^{a-1} (1-x)^{b-1} = \\
& = \frac{\Gamma(a) \Gamma(b)}{\Gamma(a+b)} \equiv \frac{a+b}{a \cdot b} \prod_{k \geq 1} \frac{\left(1 + \frac{a+b}{k}\right)}{\left(1 + \frac{a}{k}\right) \left(1 + \frac{b}{k}\right)} \quad (4.119)
\end{aligned}$$

is *Euler's integral of the first kind* (a.k.a. Euler's *beta-function*) [3, 13].

We define the function

$$I(x, y) \stackrel{\text{def}}{=} B(-\alpha'x - 1, -\alpha'y - 1) \quad (4.120)$$

so that, using the identity (4.115), the 4-tachyon disk amplitude becomes

$$2i g_0^4 C (2\pi)^{26} \delta^{26}(\Sigma_i k_i) \left[I(s, u) + I(t, u) + I(t, s) \right]. \quad (4.121)$$

The expression (4.121) is the *Veneziano amplitude* introduced before the birth of string theory to model hadron phenomenology. For the historical motivation which led Veneziano to his celebrated amplitude, see, for example, [14].

Veneziano Amplitude: Poles and Unitarity From the RHS of Eq. (4.119) we see that, as a function of s , the amplitude (4.121) has simple poles at

$$\alpha' s = -1, 0, 1, 2, \dots, n, \dots \quad (4.122)$$

which are the mass-squares of the open string states, and no other singularities. The positions of the poles in t and u are the same.

By unitarity,⁷ the residue of the tachyonic pole at $s = -1/\alpha'$ should be the square of the tree-level 3-tachyon amplitude. This determines the overall normalization constant C in term of the constant in the normalization of the operators:

Exercise 4.2 Show that $C = 1/(\alpha' g_0^2)$.

The residue of poles at $\alpha' s$ an even integer cancels between the 3 terms in (4.121). We leave to check this as an **exercise**.

⁷ Of course, the Veneziano amplitude, being a tree-level amplitude, is not fully unitary (in particular it singularities are just poles and the unitary cuts are not present. Here “unitarity” is used in the tree-level sense, that is, the residues of poles should be positive.

BOX 4.2 - Properties of the 2-2 S -matrix

We consider the scattering of two particles with momenta k_1, k_2 into two particles of momenta $-k_3, -k_4$. We use the Mandelstam variables s, t, u

$$s + t + u = \sum_i m_i^2.$$

For 4 particles of equal mass m , the scattering angle in the center of mass frame is

$$\cos \theta = 1 + \frac{2t}{s - 4m^2}.$$

In d spacetime dimensions, the residue of the S -matrix at a pole at $s = s_0$ must have the form

$$\sum_{\ell} f_{\ell} C_{\ell}^{(\lambda)}(\cos \theta) = \sum_{\ell} f_{\ell} C_{\ell}^{(\lambda)} \left(1 + \frac{2t}{s_0 - 4m^2} \right), \quad \lambda \equiv \frac{d-3}{2} \quad (\diamond)$$

where $C_{\ell}^{(\lambda)}$ ($\ell = 0, 1, 2, \dots$) are ultraspherical (Gegenbauer) polynomials [3] of degree ℓ , which generalize the Legendre ones $P_{\ell}(\cos \theta)$ for $d = 4$. Their generating function is

$$(1 - 2xz + z^2)^{-\lambda} = \sum_{\ell=0}^{\infty} z^{\ell} C_{\ell}^{(\lambda)}(x).$$

f_{ℓ} is the square of the amplitude to produce single particles of mass-square s_0 and spin ℓ , so by *unitarity* it must be non-negative. The residue at $s = s_0$ is a polynomial $P_{s_0}(t)$ in t of degree $\ell_{\max}$, where $\ell_{\max}$ is the maximal spin of a particle of mass-square s_0 . We note the formula

$$\begin{aligned} C_{\ell}^{(\lambda)}(\cos \theta) &= \sum_{k=0}^{[\ell/2]} \frac{(-1)^k (\lambda)_{\ell-k}}{k! (\ell-2k)!} (2 \cos \theta)^{\ell-2k} = \\ &= \frac{2^{\ell-2} (\lambda)_{\ell-1}}{(\ell-2)!} \left[\frac{4(\lambda + \ell - 1)}{\ell(\ell-1)} (\cos \theta)^{\ell} - (\cos \theta)^{\ell-2} + \dots \right], \end{aligned} \quad (\spadesuit)$$

where $(a)_n = a(a+1)(a+2) \cdots (a+n-1)$ is Pochhammer's symbol.

Claim. Let $A(s, t)$ be a 2-2 S -matrix amplitude (all particles of equal mass-square m^2) which has a pole at $s = s_0$ corresponding to a particle of spin $\ell > 1$. Let $P_{s_0}(\cos \theta)$ be the residue at $s = s_0$ which is a polynomial of degree ℓ . Unless all coefficients of $P_{s_0}(\cos \theta)$ are non-negative, there is a maximal dimension $d_{\max} < \infty$ in which $A(s, t)$ is consistent with unitarity. Indeed

$$\begin{aligned} P_{s_0}(\cos \theta) &= a (\cos \theta)^{\ell} + b (\cos \theta)^{\ell-1} + c (\cos \theta)^{\ell-2} + \dots = \\ &= f_{\ell} C_{\ell}^{(\lambda)}(\cos \theta) + f_{\ell-1} C_{\ell-1}^{(\lambda)}(\cos \theta) + f_{\ell-2} C_{\ell-2}^{(\lambda)}(\cos \theta) + \dots \end{aligned}$$

Since $f_k \geq 0$, the ratio a/c is not larger than the ratio of the coefficients of $(\cos \theta)^{\ell}$ and $(\cos \theta)^{\ell-2}$ in $C_{\ell}^{(\lambda)}(\cos \theta)$; from $(\spadesuit)$, we get the unitarity bound

$$\frac{a}{c} \leq -4 \frac{\lambda + \ell - 1}{\ell(\ell-1)} \Rightarrow d \leq -\frac{a}{2c} \ell(\ell-1) + 5 - 2\ell. \quad (\clubsuit)$$

BOX 4.3 - Veneziano amplitude versus unitarity and critical dimension

The Veneziano amplitude has poles at

$$\alpha' s + 1 = n = 0, 1, 2, \dots,$$

with residue polynomial of degree $\ell \equiv n$

$$\begin{aligned} P_n(t) &= \prod_{k=1}^n (t + 1 + k) = 2^{-n} \prod_{k=1}^n [(n+3) \cos \theta + 2k - n - 1] = \\ &= 2^{-n} \left\{ (n+3)^n (\cos \theta)^n - \frac{(n^2 - 1)n}{6} (n+3)^{n-2} (\cos \theta)^{n-2} + \dots \right\} \end{aligned}$$

where we used $t = (n+3)(\cos \theta - 1)/2$ and the identity

$$\sum_{1 \leq k < h \leq n} (2k - n - 1)(2h - n - 1) = -\frac{(n^2 - 1)n}{6}.$$

Comparing with the **Claim** in the **BOX 4.2**, we see that

$$-\frac{a_n}{c_n} = \frac{6(n+3)^2}{(n^2 - 1)n}$$

and the unitarity bound at the n -th pole gives

$$d \leq d_n \equiv \frac{3(n+3)^2}{n+1} + 5 - 2n$$

where $\{d_0, d_1, d_2, d_3, d_4, d_5\} = \{32, 27, 26, 26, 26 + \frac{2}{5}, 27\}$ and $d_n > 27$ for $n > 5$. We see that $d = 26$ is the largest possible dimension in which the Veneziano amplitude is consistent with unitarity (it saturates the bound for $n = 2, 3$).

Veneziano Amplitude: High-Energy Behavior

There are two regions of interest

- the *Regge limit*

$$s \rightarrow \infty, \quad t \text{ fixed}, \quad (4.123)$$

- the *hard scattering limit*

$$s \rightarrow \infty, \quad t/s \text{ fixed}. \quad (4.124)$$

Regge Limit For large x one has the Stirling asymptotic expansion

$$\Gamma(x+1) \approx x^x e^{-x} \sqrt{2\pi x} \left(1 + \sum_{k \geq 1} \frac{g_k}{x^k}\right), \quad (4.125)$$

where $g_1 = 1/12$, $g_2 = 1/288$, *etc.*; see, for example, [3, 13].⁸ Then in the Regge limit the amplitude is proportional to

$$s^{\alpha' t + 1} \Gamma(-\alpha' t - 1), \quad (4.126)$$

that is, it varies as a power of s with a t -dependent exponent. At the poles of the amplitude the exponent is a non-negative integer ℓ corresponding to the exchange of particles with maximal spin ℓ (see **BOX 4.2** for general properties of scattering matrices, including the relation of the exponent with spin and unitarity constraints). Hence for high masses we have the behavior

$$\text{maximal spin of a particle of mass } m = \alpha' m^2 + 1, \quad (4.127)$$

that is, spin is linear in the mass-square (this is the so-called *Regge behavior*) with slope⁹ α' and *intercept* 1.

In **BOX 4.3**, we give a new interpretation of the critical dimension $d = 26$ of the bosonic string: we show that $d = 26$ is the *maximal* spacetime dimension in which the Veneziano amplitude is consistent with tree-level unitarity.

Veneziano Amplitude: Hard Scattering Limit Let us go to the hard scattering regime. From formulae in **BOX 4.2**, we see that in the limit (4.124)

$$\begin{aligned} 1 - 2 \sin^2(\theta/2) &\equiv \cos \theta = 1 + \frac{2t}{s}, \\ \Rightarrow \frac{t}{s} &= -\sin^2(\theta/2), \quad \frac{u}{s} = -\cos^2(\theta/2). \end{aligned} \quad (4.128)$$

The amplitude goes like

$$\exp\left[-\alpha'(s \log(s\alpha') + t \log(t\alpha') + u \log(u\alpha'))\right] = \exp\left[-\alpha' s f(\theta)\right] \quad (4.129)$$

where

$$f(\theta) = -\sin^2(\theta/2) \log \sin^2(\theta/2) - \cos^2(\theta/2) \log \cos^2(\theta/2) > 0. \quad (4.130)$$

In QFT hard scattering at fixed angle falls as a power of s . The exponential fall-off is much softer: it suggests a smooth object of size $\sqrt{\alpha'}$, as expected.

⁸ For an *exact* version of the Stirling formula, see **Theorem 1.4.2** of [13].

⁹ This fact explains why α' is also called the *Regge slope*.

4.4 Chan–Paton Labels and Gauge Interactions

In Sect. 3.9 we introduced the Chan–Paton labels for the open bosonic string and discussed how these d.o.f. lead to non-Abelian gauge interactions. Here we check that prediction by computing tree-level amplitudes and comparing with the corresponding quantities in Yang–Mills theory. The only effect of the CP d.o.f. is to introduce new group-theoretic factors in the amplitudes according to the rules of Sect. 3.9.

3-Tachyon Disk Amplitude The 3-tachyon disk amplitude is

$$\frac{ig_0}{\alpha'} (2\pi)^{26} \delta(\Sigma_i k_i) \text{tr}(\lambda^{a_1} \lambda^{a_2} \lambda^{a_3} + \lambda^{a_1} \lambda^{a_3} \lambda^{a_2}). \quad (4.131)$$

Note that the two cyclic orderings of the vertices have different Chan–Paton traces.

4-Tachyon Disk Amplitude The contribution to the Veneziano amplitude from each cyclic order may be read in Eq. (4.118). Then the 4-tachyon disk amplitude is

$$\begin{aligned} \frac{ig_0^2}{\alpha'} (2\pi)^{26} \delta(\Sigma_i k_i) & \left[\text{tr}(\lambda^{a_1} \lambda^{a_2} \lambda^{a_4} \lambda^{a_3} + \lambda^{a_1} \lambda^{a_3} \lambda^{a_4} \lambda^{a_2}) B(-\alpha(s), -\alpha(t)) + \right. \\ & + \text{tr}(\lambda^{a_1} \lambda^{a_3} \lambda^{a_2} \lambda^{a_4} + \lambda^{a_1} \lambda^{a_4} \lambda^{a_2} \lambda^{a_3}) B(-\alpha(t), -\alpha(u)) + \\ & \left. + \text{tr}(\lambda^{a_1} \lambda^{a_2} \lambda^{a_3} \lambda^{a_4} + \lambda^{a_1} \lambda^{a_4} \lambda^{a_3} \lambda^{a_2}) B(-\alpha(s), -\alpha(u)) \right] \end{aligned} \quad (4.132)$$

where we set $\alpha(s) \equiv \alpha' s + 1$.

Exercise 4.3 Check that the unitarity relation between the residue of the pole at $s = -\alpha'$ and the square of the tree-level 3-tachyon amplitude is still valid.

Gauge Interactions

We write the open massless vector vertex of polarization e_μ and momentum k as

$$-g'_0 e_\mu \star c \dot{X}^\mu e^{ik \cdot X} \star \quad (4.133)$$

where g'_0 is an overall normalization constant. The amplitude for a gauge boson of polarization $e_{1\mu}$ and momentum k_1 and two tachyons is

$$\begin{aligned} & -ig'_0 g_0^2 e^{-\lambda} e_{1\mu} \left\langle \star c \dot{X}^\mu e^{ik_1 X}(x_1) \star \star c e^{ik_2 X}(x_2) \star \star c e^{ik_3 X}(x_3) \star \right\rangle \times \\ & \times \text{tr}(\lambda^{a_1} \lambda^{a_2} \lambda^{a_3}) + (k_2, a_2) \leftrightarrow (k_3, a_3). \end{aligned} \quad (4.134)$$

Using the formula (4.32) for correlators in the disk we have

$$\begin{aligned} & \left\langle \star c \dot{X}^\mu e^{ik_1 X}(x_1) \star \star c e^{ik_2 X}(x_2) \star \star c e^{ik_3 X}(x_3) \star \right\rangle_{\text{disk}} = -2i\alpha' \left(\frac{k_2^\mu}{x_{12}} + \frac{k_3^\mu}{x_{13}} \right) \times \\ & \times iC(2\pi)^{26} \delta^{26}(\Sigma_i k_i) |x_{12}|^{2\alpha' k_1 k_2} |x_{13}|^{2\alpha' k_1 k_3} |x_{23}|^{2\alpha' k_2 k_3}. \end{aligned} \quad (4.135)$$

Again the dependence on the x_i 's is canceled by the ghost correlator. We may simplify the expression using momentum conservation, the mass-shell conditions, and transversality $k_1 \cdot e_1 = 0$.

Exercise 4.4 Show that the amplitude may be written in the form ($k_{ij} \equiv k_i - k_j$)

$$-i g'_0 (2\pi)^{26} \delta^{26}(\Sigma_i k_i) e_1 \cdot k_{23} \operatorname{tr}(\lambda^{a_1} [\lambda^{a_2}, \lambda^{a_3}]). \quad (4.136)$$

In the presence of Chan–Paton d.o.f., the residue at $s = 0$ of the 4-tachyon amplitude does not vanish since the various terms have different Chan–Paton factors. We get a residue proportional to

$$\operatorname{tr}([\lambda^{a_1}, \lambda^{a_2}][\lambda^{a_3}, \lambda^{a_4}]). \quad (4.137)$$

Exercise 4.5 Relate the residue at $s = 0$ of the 4-tachyon amplitude with the 2-tachyon/one massless vector amplitude by unitarity and show that $g'_0 = (2\alpha')^{-1/2} g_0$. Check that the relative normalization of the tachyon and vector vertices is consistent with the state/operator correspondence.

Exercise 4.6 Show that the 3 massless vector amplitude is

$$i g'_0 (2\pi)^{26} \delta^{26}(\Sigma_i k_i) \operatorname{tr}(\lambda^{a_1} [\lambda^{a_2}, \lambda^{a_3}]) T^{\mu\nu\rho} e_{1\mu} e_{2\nu} e_{3\rho} \quad (4.138)$$

where $T^{\mu\nu\rho}$ is the tensor

$$T^{\mu\nu\rho} = k_{23}^\mu \eta^{\nu\rho} + k_{31}^\nu \eta^{\rho\mu} + k_{12}^\rho \eta^{\mu\nu} + \frac{\alpha'}{2} k_{23}^\mu k_{31}^\nu k_{12}^\rho. \quad (4.139)$$

Physical Discussion To the first order in momenta, all the above amplitudes are reproduced by the spacetime action

$$\frac{1}{g_0'^2} \int d^{26}x \left[-\frac{1}{2} \operatorname{tr}(D_\mu \phi D^\mu \phi) + \frac{1}{2\alpha'} \operatorname{tr} \phi^2 + \frac{\sqrt{2}}{3\sqrt{\alpha'}} \operatorname{tr} \phi^3 - \frac{1}{4} \operatorname{tr} F_{\mu\nu} F^{\mu\nu} \right] \quad (4.140)$$

where the tachyon ϕ and the Yang–Mills vector field A_μ are written as $n \times n$ matrices

$$\phi = \phi^a \lambda_{ij}^a, \quad A_\mu = A_\mu^a \lambda_{ij}^a, \quad (4.141)$$

and D_μ is the covariant derivative

$$D_\mu \phi = \partial_\mu \phi - i[A_\mu, \phi] \quad (4.142)$$

while

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu]. \quad (4.143)$$

Equation (4.140) is the action of $U(n)$ YM coupled to a (tachyonic) adjoint scalar. Introducing the Chan–Paton d.o.f. yields non-Abelian gauge-invariant interactions. Gauge invariance is guaranteed a priori since the unphysical states decouple by the string no-ghost theorem.

The term of order $\alpha' k^3$ in the 3 vector amplitude (4.138) implies the presence of a higher derivative term in the spacetime action

$$\frac{-2i\alpha'}{3g'^2} \text{tr}(F_{\mu\nu} F^\nu{}_\rho F^{\rho\nu}). \quad (4.144)$$

From the 4 vector amplitude we get an infinite series of higher order interactions besides this one. All perturbative states in the open strings transform in the adjoint of $U(n) = U(1) \times SU(n)$; in particular they are inert for the Abelian part $U(1)$.

We note that the $U(n)$ symmetry acting on the Chan–Paton world-sheet d.o.f., which is a *global* symmetry at the world-sheet level, is promoted to a *local* symmetry in spacetime. This is a general feature: global symmetries on the world-sheet become gauge symmetries in spacetime. Thus in (super)string theory the physics in spacetime cannot have any *global* symmetry. This is a very *deep* fact which holds in all consistent theories containing Quantum Gravity; see [15, 16] for reviews.

4.5 Closed String Tree-Level Amplitudes

3 Closed-String Tachyons

The sphere amplitude for three closed string tachyons is¹⁰

$$g_c^3 e^{-2\lambda} \left\langle \prod_{i=1}^3 : \tilde{c} c e^{ik_i X}(z_i, \bar{z}_i) : \right\rangle_{S^2} \quad (4.145)$$

where we used the $PSL(2, \mathbb{C})$ symmetry to fix the three vertices to arbitrary positions z_i (the amplitude being independent of the choices). Using the correlators computed at the beginning of this chapter, we get

$$i g_c^3 C_{S^2} (2\pi)^{26} \delta^{26}(\Sigma_i k_i) \quad (4.146)$$

where the constant C_{S^2} is $e^{-2\lambda}$ times the functional determinants (see Sects. 4.1, 4.2).

4 Closed-String Tachyons (the Virasoro–Shapiro Amplitude)

For 4 closed string tachyons the amplitude is

$$g_c^4 e^{-2\lambda} \int_{\mathbb{C}} d^2 z_4 \left\langle \prod_{i=1}^3 : \tilde{c} c e^{ik_i X}(z_i, \bar{z}_i) : : e^{ik_4 X}(z_4, \bar{z}_4) : \right\rangle_{S^2}. \quad (4.147)$$

¹⁰ g_c is the normalization constant for the closed string tachyon vertex.

For convenience we fix the first 3 vertices at 0, 1, and ∞ , respectively. The resulting amplitude is

$$i g_c^4 C_{S^2} (2\pi)^{26} \delta^{26}(\Sigma_i k_i) J(s, t, u) \quad (4.148)$$

where we defined the function

$$J(s, t, u) = \int_{\mathbb{C}} d^2 z |z|^{-\alpha' u/2-4} |1-z|^{-\alpha' t/2-4} \quad \text{with } s+t+u \equiv -\frac{16}{\alpha'} \quad (4.149)$$

which is symmetric in its 3 variables. One has the integral formula (see **BOX 4.4**)

$$C(a, b) \stackrel{\text{def}}{=} \int_{\mathbb{C}} d^2 z |z|^{2a-2} |1-z|^{2b-2} = 2\pi \frac{\Gamma(a) \Gamma(b) \Gamma(c)}{\Gamma(a+b) \Gamma(a+c) \Gamma(b+c)} \quad (4.150)$$

where $a+b+c=1$. Hence the 4 closed string tachyon amplitude is

$$i g_c^4 C_{S^2} (2\pi)^{26} \delta^{26}(\Sigma_i k_i) C(-1-\alpha' t/4, -1-\alpha' u/4) \quad (4.151)$$

known as the *Virasoro–Shapiro amplitude*. As the Veneziano amplitude, it was discovered before the introduction of string theory when people were looking for S -matrix tree-level amplitudes with “magic” properties. From its expression as a product of Gamma functions, we see that its poles are at

$$\alpha' s, \alpha' t, \alpha' u = -4, 0, 4, 8, 12, \dots \quad (4.152)$$

which are the mass-squared of the closed string states. The pole at $\alpha' s = -4$ is

$$-\frac{8\pi i g_c^3 C_{S^2}}{\alpha' s + 4} + \text{regular around } s = -4/\alpha' \quad (4.153)$$

more and then unitarity yields

$$C_{S^2} = \frac{8\pi}{\alpha' g_c^2}. \quad (4.154)$$

Exercise 4.7 Show that the Virasoro–Shapiro amplitude has Regge behavior in the Regge limit and exponential behavior in the hard scattering limit.

1 Massless Boson, 2 Tachyons

The amplitude on the sphere S^2 for 2 closed string tachyons and one closed string massless boson (graviton, B -field, or dilaton) is

BOX 4.4 - Computation of the Virasoro–Shapiro integral

We want to compute

$$C(a, b) = \int_{\mathbb{C}} d^2z |z|^{2a-2} |1-z|^{2b-2}$$

where $d^2z = i dz \wedge d\bar{z} \equiv 2 dx \wedge dy$. We use the elementary identity

$$|z|^{2a-2} = \frac{1}{\Gamma(1-a)} \int_0^\infty dt t^{-a} e^{-|z|^2 t}$$

to rewrite

$$C(a, b) = \frac{1}{\Gamma(1-a)\Gamma(1-b)} \int_0^\infty du dt t^{-a} u^{-b} \int_{\mathbb{C}} d^2z e^{-|z|^2 t - |1-z|^2 u}.$$

The integral in z is now Gaussian, and we get

$$C(a, b) = \frac{2\pi}{\Gamma(1-a)\Gamma(1-b)} \int_0^\infty du dt \frac{t^{-a} u^{-b}}{t+u} e^{-tu/(t+u)}.$$

We change variables setting $\alpha = t + u$ with $\alpha \in [0, \infty)$ and $t = \beta(t + u)$, $u = (1 - \beta)(t + u)$ with $\beta \in [0, 1]$. One has $du \wedge dt = \alpha d\alpha \wedge d\beta$. Then

$$\begin{aligned} C(a, b) &= \frac{2\pi}{\Gamma(1-a)\Gamma(1-b)} \int_0^1 d\beta \int_0^\infty d\alpha \alpha^{-a-b} \beta^{-a} (1-\beta)^{-b} e^{-\beta(1-\beta)\alpha} \\ &= 2\pi \frac{\Gamma(1-a-b)}{\Gamma(1-a)\Gamma(1-b)} \int_0^1 d\beta \beta^{b-1} (1-\beta)^{a-1} \\ &= 2\pi \frac{\Gamma(1-a-b)\Gamma(a)\Gamma(b)}{\Gamma(1-a)\Gamma(1-b)\Gamma(a+b)} \end{aligned}$$

which is the formula to be shown. In the last two steps, we used Euler's 1st and 2nd integral [3].

$$\begin{aligned} g_c^2 g_c' e^{-2\lambda} e_{\mu\nu} \left\langle : \tilde{c} c \partial X^\mu \bar{\partial} X^\nu e^{ik_1 X}(0) : : \tilde{c} c e^{ik_2 X}(1) : : \tilde{c} c e^{ik_3 X}(\infty) : \right\rangle_{S^2} = \\ = -\frac{\pi i \alpha'}{2} g_c' (2\pi)^{26} \delta^{26}(\Sigma_i k_i) e_{\mu\nu} k_{23}^\mu k_{23}^\nu \end{aligned} \quad (4.155)$$

where $e_{\mu\nu}$ is the polarization of the massless boson normalized as $e_{\mu\nu} e^{\mu\nu} = 1$. To get the amplitude involving a graviton, B -field, or dilaton, one simply specializes the above formula to, respectively, $e_{\mu\nu}$ symmetric traceless, antisymmetric, or pure trace. Since the amplitude satisfies the Ward identities of local reparametrization and B -field gauge symmetry, it must coincide with the unique one¹¹ following from the universal general covariant a gauge-invariant coupling of the tachyon to the metric,

¹¹ To the leading order in momenta.

B -field, and dilaton. Expanding the Virasoro–Shapiro amplitude around the pole at $s = 0$, we see that unitarity requires

$$g_c = \frac{2}{\alpha'} g_c \quad (4.156)$$

in agreement with the CFT state-operator correspondence.

3 Massless Bosons

Likewise the sphere amplitude for 3 massless closed-string bosons is

$$\frac{i\kappa}{2} (2\pi)^{26} \delta^{26}(\Sigma_i k_i) e_{1\mu\nu} e_{2\rho\sigma} e_{3\gamma\delta} T^{\mu\rho\gamma} T^{\nu\sigma\delta} \quad (4.157)$$

where the tensor $T^{\mu\nu\rho}$ is

$$T^{\mu\rho\gamma} = k_{23}^\mu \eta^{\rho\gamma} + k_{31}^\rho \eta^{\gamma\mu} + k_{12}^\gamma \eta^{\mu\rho} + \frac{\alpha'}{8} k_{23}^\mu k_{31}^\rho k_{12}^\gamma, \quad (4.158)$$

and κ is the gravitational coupling

$$\kappa = \pi \alpha' g_c' = 2\pi g_c. \quad (4.159)$$

Again we get the amplitudes with any combination of three gravitons, B -fields, and dilatons by specializing the polarization tensors $e_{a\mu\nu}$ to definite irreducible representation of the Lorentz group.

The terms of order $O(k^2)$ in the amplitude precisely correspond to the interactions in the 2-derivative spacetime action of Sect. 1.8.1, and the terms of order $O(k^4)$ and $O(k^6)$ to higher derivative couplings which contain operators quadratic and cubic in the spacetime curvature.

Exercise 4.8 Fill in the details of the derivation of Eqs. (4.157)–(4.159).

Exercise 4.9 Check that the above amplitude coincides with the on-shell 3-point amplitudes for massless fields for the effective Lagrangian in Sect. 1.8.1.

4.5.1 Closed String Amplitudes on the Disk and $\mathbb{RP}^2$

Amplitudes of closed string vertices on the disk or $\mathbb{RP}^2$ are of order $O(e^{-\lambda})$. Since closed string g -loop amplitudes scale as $e^{-2(1-g)\lambda}$, they are “half-loop” order.

The one massless closed-string boson amplitude on the sphere vanishes by conformal invariance (cf. Eq. (2.180)). This is *not* true for the disk or $\mathbb{RP}^2$. The one-point amplitude is necessarily at zero-momentum; then the amplitude is proportional to

$$-\left\langle : \partial X^\mu \bar{\partial} X^\nu : \right\rangle_{\text{disk}} = C \eta^{\mu\nu} \quad (4.160)$$

where the correlation function is computed from the scalars' path integral. C is a non-zero constant

$$C = \langle 1 \rangle_{S^2} \frac{\alpha'}{2}. \quad (4.161)$$

Indeed the image method relates this disk amplitude to the sphere one with ∂X^μ inserted at the two poles; see Eq. (4.29).

A one-point amplitude is known as a *tadpole*. The tadpole (4.160) corresponds to a spacetime interaction of the form

$$\text{const.} \int d^{26}x \sqrt{-G} e^{-\Phi} \quad (4.162)$$

which is a potential for the dilaton arising at “ $\frac{1}{2}$ -loop” order. The fact that the quantum corrections produce a non-trivial potential for Φ means that the flat vacuum we are expanding around is not perturbatively stable. All tadpoles should vanish if the vacuum is stable in perturbation theory, so the presence of non-zero tadpoles implies an inconsistency of the perturbative formulation. In other words: *we must impose absence of tadpoles as a requirement of the perturbative theory.*

The above computation shows that open oriented bosonic strings are plagued by tadpoles and so are **not** fully consistent in perturbation theory around flat space.

The same observation applies for the one-point amplitude on $\mathbb{RP}^2$, again as a consequence of the image trick. This is also a “half-loop” contribution since $\chi(\mathbb{RP}^2) = 1$.

Exercise 4.10 Compute the massless boson tadpole on $\mathbb{RP}^2$.

In the open non-oriented bosonic string, the tadpole of a massless closed string state at “half-loop order” gets two contributions: one from the disk and one from the $\mathbb{RP}^2$ amplitude. The first amplitude is proportional to n because of the trace over the Chan–Paton index associated with the boundary $\partial D = S^1$. Instead the amplitude on $\mathbb{RP}^2$ is independent of n because this surface has no boundary. The amplitude carries an overall sign which depends on whether the Ω -projection leads to $SO(n)$ or $Sp(k)$.

Exercise 4.11 Show that the “half-loop” tadpole cancels for $G = SO(2^{13})$.

4.6 One-Loop Amplitudes: The Torus

We see the torus as the quotient of $\mathbb{C}$ by the lattice $\Lambda = 2\pi\mathbb{Z} \oplus 2\pi\tau\mathbb{Z}$ with $\tau \equiv \tau_1 + i\tau_2$, the *modulus* (or *period*) of the torus which is a point in the upper half-plane

$$\mathcal{H} = \{\tau \in \mathbb{C} : \text{Im } \tau > 0\}. \quad (4.163)$$

τ is a local coordinate in the complex moduli space $\mathcal{M}_{1,0}$ of genus one curves without punctures. We may write the complex coordinate z on $\mathbb{C}$ as $z = x + \tau y$, where x, y

are real coordinates of period 2π . The variation of the flat Kähler metric $ds^2 = dz d\bar{z}$ with τ is

$$\begin{aligned} \delta(dz d\bar{z}) &= \delta\tau dy d\bar{z} + \delta\bar{\tau} dz dy = \delta\tau \frac{dz - d\bar{z}}{\tau - \bar{\tau}} d\bar{z} + \delta\bar{\tau} dz \frac{dz - d\bar{z}}{\tau - \bar{\tau}} = \\ &= \frac{\delta\tau_2}{\tau_2} dz d\bar{z} + i \frac{\delta\tau}{2\tau_2} d\bar{z}^2 - i \frac{\delta\bar{\tau}}{2\tau_2} dz^2. \end{aligned} \quad (4.164)$$

On the torus there is one complex CKV, $\partial_{\bar{z}}$, which generates translation automorphism on the torus,¹² and one complex modulus τ , associated with the (conjugate) quadratic differential $d\bar{z}^2$; see (4.164). Then we have one zero-mode for each ghost field $c(z)$ and $b(z)$ and the zero-modes are translational invariant. The zero-mode of $b(z)$ describes an infinitesimal deformation of the complex structure, and so yields a 1-form on $\mathcal{M}_{1,0}$. The appropriately normalized operator insertion of the b -ghost zero-mode is then

$$\frac{1}{2\pi} \int d^2z b_{z\bar{z}}(z) \partial_{\tau} g_{z\bar{z}} d\tau = \frac{i}{4\pi\tau_2} \int d^2z b(z) d\tau \rightsquigarrow 2\pi i b(0) d\tau \quad (4.165)$$

where we used Eq. (4.164) and (in the last step) translational invariance of the zero-mode together with the volume of the torus

$$\text{vol}(T^2) = i \int_{T^2} dz \wedge d\bar{z} = 8\pi^2 \tau_2. \quad (4.166)$$

For the normalization of the c -ghost zero-mode we may average over the group of complex automorphisms—which, for a generic torus,¹³ is isomorphic to $T^2 \rtimes \mathbb{Z}_2$ —and then divide by its volume which is twice the volume of the torus, Eq. (4.166).

For the amplitude then we get

$$\int_{F_0} \frac{d\tau d\bar{\tau}}{4\tau_2} \left\langle b(0) \tilde{b}(0) \tilde{c}(0) c(0) \prod_{i=1}^n \int d\bar{z}_i d\bar{z}_i V_i(z, \bar{z}) \right\rangle_{T^2} \quad (4.167)$$

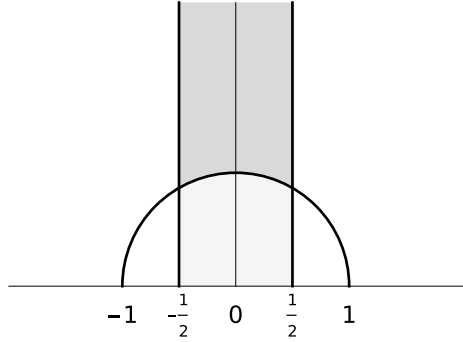
where V_i are the integrated version of the physical (BRST-invariant) vertices associated with the **in**- and **out**- asymptotic states. In Eq. (4.167) F_0 stands for a *fundamental domain* of the action of $PSL(2, \mathbb{Z})$ on the upper half-plane $\mathcal{H}$. We recall the definition

Definition 4.1 A *fundamental domain* for a group $G \subset SL(2, \mathbb{R})$ acting on $\mathcal{H}$ is a connected open set $F \subset \mathcal{H}$ bounded by smooth arcs $\partial\bar{F} = \cup_i (c_i^+ \cup c_i^-)$ such that

¹² The torus is an Abelian group, and the translation automorphism is induced by its group structure.

¹³ As we shall see, the extra automorphism for special values of the modulus τ are automatically taken care since these special points correspond to orbifold points of $\mathcal{M}_{1,0} \equiv \mathcal{H}/PSL(2, \mathbb{Z})$ associated with the torsion subgroups of $PSL(2, \mathbb{Z})$.

Fig. 4.3 The *fundamental domain* F_0 for the action on $SL(2, \mathbb{Z})$ on the upper half-plane $\mathcal{H}$ (darker gray region). The half-strip R in Eq. (4.179) is the part of the upper half-plane between the two vertical lines $\text{Re } \tau = -1/2$ and $\text{Re } \tau = 1/2$ (union of dark and light gray regions)



1. for each $w \in \mathcal{H}$ there is a $z \in \overline{F}$ and a $g \in G$ such that $g(w) = z$;
2. if $z_1, z_2 \in F$ and $g(z_1) = z_2$ for $g \in G$ then $g = 1$;
3. for each i there exists $g_i \in G$ with $g_i(c_i^+) = c_i^-$.

Integrating over F_0 is obviously equivalent to integrating over

$$\mathcal{M}_{1,0} \equiv \mathcal{H} / PSL(2, \mathbb{Z}), \quad (4.168)$$

i.e. over the inequivalent conformal structures on the torus. In other words, to avoid double counting, we have to restrict to a fundamental domain of the action of $PSL(2, \mathbb{Z})$ on the upper half-plane. A convenient choice is the domain

$$\overline{F}_0 = \left\{ \tau \in \mathcal{H} : |\text{Re } \tau| \leq \frac{1}{2}, |\tau|^2 \geq 1 \right\}; \quad (4.169)$$

see **Fig. 4.3**.

The Vacuum Amplitude We consider first the case of no insertion

$$Z_{T^2} = \int_{F_0} \frac{d\tau d\bar{\tau}}{4\tau_2} \left\langle b(0) \tilde{b}(0) \tilde{c}(0) c(0) \right\rangle_{T^2}. \quad (4.170)$$

The integrand is the product of the path integrals on the torus of period τ for the matter sector and the ghosts; see Eqs. (4.55) and (4.101). In 26 flat dimensions we get

$$\begin{aligned} Z_{T^2} &= i V_{26} \int_{F_0} \frac{d\tau d\bar{\tau}}{4\tau_2} Z_X(\tau, \bar{\tau})^{26} |\eta(\tau)|^4 \equiv \\ &\equiv i V_{26} \int_{F_0} \frac{d\tau d\bar{\tau}}{4\tau_2} (4\pi^2 \alpha' \tau_2)^{-13} |\eta(\tau)|^{-48}. \end{aligned} \quad (4.171)$$

We stress again that the net effect of the c, b -ghosts is to cancel the contribution from the oscillators of two scalars (the longitudinal ones X^+, X^-) leaving only their zero-mode part which for each non-compact scalar yields a factor $(4\pi^2 \alpha' \tau_2)^{-1/2}$ in

the integrand from the integral over the momentum in the non-compact direction. This remains true if the matter part of the world-sheet theory is an arbitrary $c = 26$ CFT as long as our target-space contains $d \geq 2$ non-compact flat directions.

The integrand in Eq. (4.171) is $SL(2, \mathbb{Z})$ modular invariant. Indeed we know from Eq. (4.58) that $\tau_2 |\eta(\tau)|^4 \equiv Z_X^{-2}$ is modular invariant; then the integrand has the form

$$\int_{F_0} \omega_P \left(\text{modular invariant function} \right), \quad \text{where } \omega_P \equiv \frac{i d\tau \wedge d\bar{\tau}}{\tau_2^2} \quad (4.172)$$

ω_P is the Kähler form of the $SL(2, \mathbb{R})$ -invariant Poincaré metric on the upper half-plane, i.e. the $SL(2, \mathbb{R})$ -invariant volume form (see **BOX 4.5**).

We can easily generalize the above analysis to the situation where the world-sheet matter CFT has the Lagrangian $\partial_a X^+ \partial^a X^- + L^\perp$, with $L^\perp$ the Lagrangian of a transverse CFT with $c = 24$. The torus partition function can be written as a trace over the Hilbert space of the transverse CFT, so the vacuum amplitudes becomes

$$Z_{T^2} = i V_d \int_{F_0} \frac{d\tau d\bar{\tau}}{4 \tau_2} (4\pi^2 \alpha' \tau_2)^{-d/2} \sum_{i \in \mathbf{H}^\perp} q^{h_i-1} \bar{q}^{\tilde{h}_i-1} \quad (4.173)$$

where $d \geq 2$ is the number of non-compact free scalars and $\mathbf{H}^\perp$ is the Hilbert space of the transverse d.o.f. with zero-modes of the non-compact scalars removed.

Physical Interpretation Let us compare this expression with the one-loop correction to the vacuum amplitude in a QFT where we sum over all particle periodic paths¹⁴

$$\begin{aligned} Z(m^2)_{\text{particle}} &= V_d \int \frac{d^d k}{(2\pi)^d} \int_0^\infty \frac{dl}{2l} \exp\left[-\frac{1}{2}(k^2 + m^2)l\right] = \\ &= i V_d \int_0^\infty \frac{dl}{2l} (2\pi l)^{-d/2} \exp\left(-\frac{1}{2}m^2 l\right); \end{aligned} \quad (4.174)$$

see **BOX 4.6**. Now we take this point-particle formula¹⁵ and sum over the particle spectrum of the bosonic string. As we have seen several times (e.g. by light-cone quantization), the string particle spectrum is in one-to-one correspondence with the states in $\mathbf{H}^\perp$ where the mass is related to the weights of the transverse CFT as

$$m^2 = \frac{2}{\alpha'} (h + \tilde{h} - 2) \quad \text{with the matching constraint } h = \tilde{h}. \quad (4.175)$$

¹⁴ m is the mass of the particle. The factor $1/2$ may be thought of as the result of modding out the reversal of the world-line coordinate.

¹⁵ The above expression for the one-loop amplitude in QFT is called the Coleman–Weinberg formula [17].

BOX 4.5 - Poincaré geometry of the upper half-plane

The upper half-plane (and all simply connected proper sub-domains of $\mathbb{C}$) is biholomorphic with the Hermitian symmetric space $SL(2, \mathbb{R})/U(1)$. The simplest way to see this is to use the Iwasawa decomposition of $SL(2, \mathbb{R})$

Lemma (Iwasawa decomposition for $SL(2, \mathbb{R})$). *All elements of $A \in SL(2, \mathbb{R})$ may be written uniquely in the form*

$$A = \begin{pmatrix} y^{1/2} & y^{-1/2}x \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \equiv T(y, x) e^{i\theta\sigma_2} \quad (\star)$$

for some $x \in \mathbb{R}$, $y \in \mathbb{R}_{>0}$ and $\theta \in [0, 2\pi)$. The map $(x, y) \mapsto z \equiv x + iy$ is a biholomorphic equivalence between $SL(2, \mathbb{R})/SO(2)$ and the upper half-plane.

We identify the upper half-plane with the space of triangular matrices $T \equiv T(y, x)$ as in the first factor in the RHS of $(\star)$. The group $SL(2, \mathbb{R})$ acts on itself, hence on the upper half-plane, by multiplication on the left $A \mapsto BA$. The Maurer–Cartan form is

$$\omega_{\text{MC}} \stackrel{\text{def}}{=} A^{-1} dA = e^{-i\theta\sigma_2} \frac{1}{2y} \begin{pmatrix} dy & 2dx \\ 0 & -dy \end{pmatrix} e^{i\theta\sigma_2} + i\sigma_2 d\theta.$$

By Cartan's theory, the $SL(2, \mathbb{R})$ -invariant metric is given (up to overall scale) by the trace of the square of the symmetric part of ω_{MC}

$$ds_{\text{inv.}}^2 = \text{tr}(\omega_{\text{MC}} + \omega_{\text{MC}}^t)^2 = \text{tr}(A^{-1} dA + dA^t A^t)^2;$$

indeed this expression is manifestly invariant under the $SL(2, \mathbb{R})$ action given by the multiplication of A on the left by a constant matrix $B \in SL(2, \mathbb{R})$ since

$$(BA)^{-1} d(BA) = A^{-1} B^{-1} (B^{-1} dA) = A^{-1} dA.$$

Now

$$ds_{\text{inv.}}^2 = \text{tr} \left[\begin{pmatrix} dy/y & dx/y \\ dx/y & -dy/y \end{pmatrix}^2 \right] = 2 \frac{dx^2 + dy^2}{y^2} \equiv 2 \frac{dz d\bar{z}}{(\text{Im } z)^2}.$$

The overall coefficient is arbitrary; we prefer to normalize the metric so that

$$ds_{\text{Poincaré}}^2 = \frac{dz d\bar{z}}{(2 \text{Im } z)^2}.$$

This metric is Kähler with Kähler potential

$$\Phi = -\log(z - \bar{z}).$$

For a Kähler metric $g_{i\bar{j}}$, the Ricci curvature is of type (1,1) with $R_{i\bar{k}} = -\partial_i \partial_{\bar{k}} \log \det g$. For the Poincaré metric normalized as above, we get

$$R_{z\bar{z}} = -g_{z\bar{z}}.$$

The Poincaré metric is Einstein: it is the only 2d metric with constant negative curvature.

We enforce the left–right matching constraint with a Kronecker delta

$$\delta_{h,\tilde{h}} = \int_{-\frac{1}{2}}^{\frac{1}{2}} dx e^{2\pi i(h-\tilde{h})x} \quad (4.176)$$

where we used that $h - \tilde{h} = N - \tilde{N}$ is an integer. Then

$$\begin{aligned} \sum_{i \in \mathbf{H}^\perp} Z(m_i^2)_{\text{particle}} &= \\ &= i V_d \int_0^\infty \frac{dl}{2l} (2\pi l)^{-d/2} \int_{-\frac{1}{2}}^{\frac{1}{2}} dx \sum_{i \in H^\perp} \exp \left[2\pi i(h - \tilde{h})x - (h_i + \tilde{h}_i - 2)l/\alpha' \right] \equiv \\ &\equiv i V_d \int_R \frac{d\tau d\bar{\tau}}{4\tau_2} (4\pi^2 \alpha' \tau_2)^{-d/2} \sum_{i \in \mathbf{H}^\perp} q^{h_i-1} \bar{q}^{\tilde{h}_i-1} \end{aligned} \quad (4.177)$$

where

$$\tau = x + il/(2\pi\alpha') \quad (4.178)$$

and R is the *half-strip*

$$R \stackrel{\text{def}}{=} \left\{ \tau = \tau_1 + i\tau_2 \in \mathbb{C} : \tau_2 > 0, \quad |\tau_1| \leq \frac{1}{2} \right\} \subset \mathcal{H}. \quad (4.179)$$

Let us interpret this result. The particle amplitude (4.174) diverges as $l \rightarrow 0$: this is the usual UV divergence of QFT. Summing over the bosonic string states makes the divergence even worse, since all contributions have the same sign. However the expression (4.177) differs from the actual string amplitude (4.173) in *one crucial aspect*: the integrands are identical but the regions of integration are different: R versus F_0 . We see that in the string amplitude the UV divergent region is simply *absent*. R contains infinitely many images under $SL(2, \mathbb{Z})$ of F_0 ; so the UV divergence of (4.177) is due to an infinite-fold double-counting.

Another possible source of divergence is the region $\tau_2 \rightarrow \infty$ where the torus is infinitely stretched. In this region the string amplitude (4.173) has the expansion

$$i V_{26} \int \frac{d\tau_2}{2\tau_2} (4\pi^2 \alpha' \tau_2)^{-13} \left[\exp(4\pi \tau_2) + 24^2 + \dots \right]. \quad (4.180)$$

The asymptotic behavior as $\tau_2 \rightarrow \infty$ is controlled by the lightest states. The first term is exponentially divergent and is due to the tachyon; this divergence will not be present in string theories which do not have a tachyon such as the superstring after GSO projection.

The torus vacuum amplitude illustrates a general feature of string perturbation theory: *there is no UV region in moduli space which may give rise to a UV divergence.*

BOX 4.6 - Point-particle vacuum amplitude

The free energy of a free scalar of mass m is formally given by

$$\log Z(m^2) = \log(\text{Det}[-\partial^2 + m^2])^{-1/2} = -\frac{1}{2} \text{Tr} \log(-\partial^2 + m^2)$$

which is defined through ζ -function regularization

$$\log Z(m^2) = \frac{1}{2} \frac{\partial}{\partial s} \text{Tr}(-\partial^2 + m^2)^{-s} \Big|_{\text{analytically continued to } s=0}.$$

$$\begin{aligned} \text{Now } \frac{1}{2} \text{Tr}(-\partial^2 + m^2)^{-s} &= \frac{1}{\Gamma(s)} \int_0^\infty \frac{dt}{2t} t^s \text{Tr} e^{-(\partial^2 + m^2)t} = \\ &= \frac{1}{\Gamma(s)} i V_d \int_0^\infty \frac{dl}{2l} (l/2)^s \int \frac{d^d p}{(2\pi)^d} e^{-(p^2 + m^2)l/2} = \\ &= \frac{1}{\Gamma(s)} i V_d \int_0^\infty \frac{dl}{2l} (l/2)^s (2\pi l)^{-d/2} e^{-m^2 l/2}. \end{aligned}$$

If the integral with $s = 0$ was convergent (which is not), the last expression would be equal to

$$s i V_d \int_0^\infty \frac{dl}{2l} (2\pi l)^{-d/2} e^{-m^2 l/2} + O(s^2).$$

Taking the derivative with respect to s and setting $s = 0$ we get the formal expression (4.174). Performing the integral in a region in the s -plane where it is convergent, we get the analytic continuation

$$\frac{1}{2} \text{Tr}(-\partial^2 + m^2)^{-s} = i V_d \frac{1}{2} (4\pi)^{-d/2} m^{d-2s} \frac{\Gamma(s - d/2)}{\Gamma(s)}.$$

All boundaries of moduli spaces are controlled by light states, so they correspond to IR regimes. Perturbative infinities, if any, are IR divergences.¹⁶

4.7 One-Loop: The Cylinder

The Amplitude

The cylinder ($\equiv$ annulus) case is similar. The cylinder has one real modulus: we see it as the segment $[0, \pi]$ times a circle of length $2\pi t$. The integrand is again given by a trace over the transverse d.o.f. of the open string times a momentum factor from the longitudinal zero-modes

¹⁶ The statement holds to any *finite* order. The perturbative series itself is expected to be divergent.



Fig. 4.4 As $s \rightarrow \infty$ the world-sheet of the cylinder amplitude gets long and thin

$$\begin{aligned} Z_{\text{cylinder}} &= \int_0^\infty \frac{dt}{t} (8\pi^2 \alpha' t)^{-1} \text{Tr}_o[\exp(-2\pi t(L_0 - 1))] = \\ &= i V_{26} n^2 \int_0^\infty \frac{dt}{2t} (8\pi^2 \alpha' t)^{-13} \eta(it)^{-24} \end{aligned} \quad (4.181)$$

which again can be understood by comparison with the field theory formula. The factor n^2 arises from the trace over the Chan–Paton d.o.f.: in the absence of boundary operator insertions, they give a factor $n^{\#(\text{boundaries})}$ which yields n^2 for the cylinder.

Exercise 4.12 Deduce (4.181) from string first principles.

The limit $t \rightarrow \infty$ is similar to the torus case: we have an exponential IR divergence from the open string tachyon.

Divergences and Tadpoles

The limit $t \rightarrow 0$ is more interesting. The cylinder has no modular invariance to reduce the domain of integration, so the UV divergence of QFT is apparently still there. What changes in the string case is its physical interpretation: now also the $t \rightarrow 0$ divergence is an *IR effect*.

Let us see how this comes about. By the conformal transformation $w \mapsto w/t$, we replace the cylinder $[0, \pi] \times \mathbb{R}/2\pi t \mathbb{Z}$ by the cylinder $[0, \pi/t] \times \mathbb{R}/2\pi \mathbb{Z}$. We set $s = \pi/t$ and use the modular property $\eta(it) = t^{-1/2} \eta(i/t)$ (cf. (4.57)) to rewrite (4.181) as

$$Z_{\text{cylinder}} = i \frac{V_{26} n^2}{2\pi (8\pi^2 \alpha')^{13}} \int_0^\infty ds \, \eta(is/\pi)^{-24}. \quad (4.182)$$

We expand the integrand for large lengths $s \rightarrow \infty$ of the cylinder

$$\eta(is/\pi)^{-24} = e^{2s} \prod_{n \geq 1} (1 - e^{-2ns})^{-24} = e^{2s} + 24 + O(e^{-2s}). \quad (4.183)$$

This is the asymptotics which is expected from expanding in a complete set of *closed* string states. Indeed, in the limit $t \rightarrow 0$ the world-sheet becomes a long thin tube which propagates closed string states between the two small boundary loops as in **Fig. 4.4**. The leading divergence in the vacuum amplitude is from the closed string tachyon. This term can be regularized by analytic continuation

$$\int_0^\infty ds \, e^{\beta s} = -\frac{1}{\beta}. \quad (4.184)$$

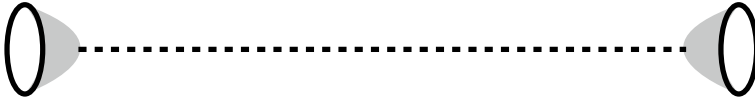


Fig. 4.5 At $s = \infty$ the cylinder amplitude describes a massless closed-string state which propagates between two disk tadpoles of its vertex

The second term, which comes from the massless closed string states, cannot be cured by analytic continuation (it corresponds to the above integral evaluated at $\beta = 0$). This bad divergence has a clear physical meaning. Consider the limit as $s \rightarrow \infty$ of the above figure: we get the situation in **Fig. 4.5**, that is, two disk tadpoles of closed string massless particles connected by a closed string propagator (the dashed line in the figure); since the massless particles have zero-momentum, the massless propagator $1/k^2$ yields an *infra-red* divergence. This IR divergence is not specific of the string: it is already present in field theory.

In QFT UV and IR divergences have quite different physical meaning. UV divergences means that at some short distance we need new physics. IR divergences often are related to non-zero tadpoles: in this case they just say that we are expanding around the wrong vacuum. If we replace it by the legitimate vacuum, the IR divergences vanish. This is what happens in bosonic string theory. The disk tadpole

$$-\Lambda \int d^{26}x \sqrt{-G} e^{-\Phi} \quad (4.185)$$

is a source for both the metric and the dilaton, and the constant backgrounds $G_{\mu\nu} = \eta_{\mu\nu}$, $\Phi = \text{const.}$ are *not* solutions of the quantum corrected equations of motion. Expanding around a solution of the correct field equations will lead to finite amplitudes. The situation in the superstring will be much better.

Open String One-Loop Scattering Amplitudes

If we insert open string vertices on the two boundaries of the cylinder $[0, \pi] \times \mathbb{R}/2\pi t\mathbb{Z}$, we get a one-loop open string scattering amplitude. In the $s \rightarrow \infty$ limit, it is more natural to see the amplitude as describing the exchange between the two boundaries of an intermediate closed string carrying a non-zero-momentum k_μ . The integrand in Eq. (4.182) includes a factor

$$e^{-\alpha' k^2 s/2} \quad (4.186)$$

and the divergence becomes a momentum pole representing scattering of open string states into closed ones.

4.8 Boundary and Cross-Cap States

Boundary States

We have introduced the cylinder $I \times S^1$ amplitude as a one-loop process for open strings by considering the circle to be periodic Euclidean time and the interval I as the coordinate along the string. We can invert the role of the two factors, and see S^1 as a closed string propagating in the Euclidean time I between two states at $\sigma^1 = 0$ and $\sigma^1 = s$. Including the ghosts, the path integral computes the matrix element

$$\langle B | c_0 b_0 \exp[-s(L_0 + \tilde{L}_0)] | B \rangle \quad (4.187)$$

where the *Neumann boundary state* $|B\rangle$ is determined by the condition that $\partial_1 X^\mu$, c^1 , and b_{12} vanish on the boundary,

$$\partial_1 X^\mu|_{\sigma^2=0}|B\rangle = c^1|_{\sigma^2=0}|B\rangle = b_{12}|_{\sigma^2=0}|B\rangle = 0, \quad (4.188)$$

that is, in terms of modes

$$(\alpha_n^\mu + \tilde{\alpha}_{-n}^\mu)|B\rangle = (c_n + \tilde{c}_{-n})|B\rangle = (b_n - \tilde{b}_{-n})|B\rangle = 0 \quad \forall n. \quad (4.189)$$

This yields [18–21][22]

$$|B\rangle = (c_0 + \tilde{c}_0) \exp \left[- \sum_{n=1}^{\infty} \left(\frac{1}{n} \alpha_{-n} \tilde{\alpha}_{-n} + b_{-n} \tilde{c}_{-n} + \tilde{b}_{-n} c_{-n} \right) \right] |0\rangle. \quad (4.190)$$

One gets the Dirichlet b.c. from the Neumann one by flipping the relative sign between left- and right-movers. The Dirichlet boundary state $|D\rangle$ then satisfies

$$(\alpha_n^\mu - \tilde{\alpha}_{-n}^\mu)|B\rangle = (c_n - \tilde{c}_{-n})|B\rangle = (b_n + \tilde{b}_{-n})|B\rangle = 0 \quad \forall n \quad (4.191)$$

and

$$|D\rangle = (c_0 - \tilde{c}_0) \exp \left[\sum_{n=1}^{\infty} \left(\frac{1}{n} \alpha_{-n} \tilde{\alpha}_{-n} + b_{-n} \tilde{c}_{-n} + \tilde{b}_{-n} c_{-n} \right) \right] |0\rangle. \quad (4.192)$$

Exercise 4.13 Check that Eqs. (4.190), (4.192), and satisfy (respectively) Eqs. (4.189), (4.191).

Cross-Cap States

$\mathbb{RP}^2 \equiv \mathbb{P}^1 / \langle \Omega \rangle$ is the quotient of the sphere by the antipodal map $\Omega: z \rightarrow -1/\bar{z}$. Writing $z = \exp[\log r + i\theta]$, the free scalar X^μ on $\mathbb{RP}^2$ lifts to a scalar on the oriented double $\mathbb{P}^1$ satisfying the reflection condition

$$X^\mu(\log r, \theta) = X^\mu(-\log r, \theta + \pi). \quad (4.193)$$

As fundamental domain for Ω we take the unit disk $\mathcal{D} = \{|z| \leq 1\}$ whose boundary we parametrize with $\theta \in [0, 2\pi]$. In the cross-cap, opposite boundary points on the unit circle $r = 1$ are identified: $\theta \sim \theta + \pi$. The cross-cap state $|C\rangle$ is defined on the unit circle $S^1 = \partial\mathcal{D}$. From Eq. (4.193)

$$\partial_\theta X^\mu(\theta)|_{r=1}|C\rangle = \partial_\theta X^\mu(\theta + \pi)|_{r=1}|C\rangle \quad (4.194)$$

$$\partial_{\log r} X^\mu(\theta)|_{r=1}|C\rangle = -\partial_{\log r} X^\mu(\theta + \pi)|_{r=1}|C\rangle \quad (4.195)$$

or, in terms of modes,

$$(\alpha_n^\mu + (-1)^n \tilde{\alpha}_{-n}^\mu)|C\rangle = 0 \quad (4.196)$$

which differs from the Neumann b.c. (4.189) only by the insertion of the sign $(-1)^n$. Hence the *matter part* of the cross-cap state $|C\rangle$ is

$$|C\rangle_{\text{matter}} = \exp\left[-\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \alpha_{-n} \tilde{\alpha}_{-n}\right]. \quad (4.197)$$

Exercise 4.14 Write the ghost factor in the cross-cap state $|C\rangle$.

4.9 One-Loop: Klein Bottle and Möbius Strip

Klein Bottle We see the Klein bottle Kl as the cylinder $[0, 2\pi t] \times \mathbb{R}/2\pi\mathbb{Z}$ where the two ends are identified with an inversion of orientation implemented by the unitary operator Ω . The amplitude for the Klein bottle then is

$$\begin{aligned} Z_{\text{Kl}} &= \int_0^\infty \frac{dt}{4t} \text{Tr}'_c[\Omega \exp(-2\pi t(L_0 + \tilde{L}_0))] = \\ &= iV_d \int_0^\infty \frac{dt}{4t} (4\pi^2 \alpha' t)^{-d/2} \sum_{i \in H_c^\perp} \Omega_i \exp[-2\pi t(h_i + \tilde{h}_i - 2)] \end{aligned} \quad (4.198)$$

where Tr'_c means omitting ghosts and longitudinal oscillators, and we have one extra factor $\frac{1}{2}$ because the projector over Ω -invariant states is

$$\frac{1}{2}(1 + \Omega) \quad (4.199)$$

such that we have to take *one-half* the sum of the torus and Kl amplitudes. We have already computed the sum for flat 26 dimensions in Sect. 4.1.1; we get



Fig. 4.6 In the crossed channel, the Klein bottle amplitude describes closed-string states propagating between two cross-cap states

$$Z_{k^2} = i V_{26} \int_0^\infty \frac{dt}{4t} (4\pi^2 \alpha' t)^{-13} \eta(2it)^{-24}. \quad (4.200)$$

The divergence as $t \rightarrow 0$ is similar to the one for the cylinder, and again has the physical interpretation of an IR singularity in terms of a massless closed string pole.

As shown in the final **example** in **BOX 1.5**, the Klein bottle of modulus t may also be interpreted as a cylinder of circumference 2π and length $s = \pi/2t$ with the two ends closed by a cross-cap (a copy of $\mathbb{RP}^2$). This dual viewpoint is more convenient, and we see **KI** as two cross-caps connected by a long cylinder; see **Fig. 4.6**. The amplitude seen from this dual perspective is obtained by a S modular transformation of the integrand in (4.200). Using

$$(2t)^{1/2} \eta(2it) = \eta(i/2t) = \eta(is/\pi) \quad (4.201)$$

we get

$$Z_{K^2} = i \frac{2^{26} V_{26}}{4\pi (8\pi^2 \alpha')^{13}} \int_0^\infty ds \eta(is/\pi)^{-24}. \quad (4.202)$$

The discussion of the leading IR divergences goes as the one for the cylinder with the massless closed-string state tadpole on the disk replaced by the corresponding tadpole on $\mathbb{RP}^2$ which is also non-zero since, by the method of images, we see that

$$\langle \partial X^\mu \bar{\partial} X^\nu \rangle_{\mathbb{RP}^2} \neq 0. \quad (4.203)$$

Exercise 4.15 Use unitarity, the cylinder amplitude, and the Klein bottle amplitude to compute the ratio between the massless tadpoles on the disk and $\mathbb{RP}^2$.

Möbius Strip We see the Möbius strip as a strip of width π and length $2\pi t$ where we identify the two ends with an orientation-reversion. Thus

$$Z_{M^2} = i V_d \int_0^\infty \frac{dt}{4t} (8\pi^2 \alpha' t)^{-d/2} \sum_{i \in H_0^\perp} \Omega_i e^{-2\pi t(h_i - 1)}. \quad (4.204)$$

As discussed in Sect. 4.1.1, the effect of Ω is to introduce an extra sign $(-1)^n$ in the sum over the oscillator occupation numbers

$$e^{2\pi t} \prod_{n \geq 1} [1 - (-1)^n e^{-2\pi n t}]^{-24} = \vartheta_3(0, 2it)^{-12} \eta(2it)^{-12}. \quad (4.205)$$

Since the Möbius strip has one boundary, we have a factor n in the amplitude from the Chan–Paton d.o.f. For the $SO(n)$ theory the $\frac{1}{2}n(n+1)$ symmetric states have $\Omega = +1$, while the $\frac{1}{2}n(n-1)$ antisymmetric states have $\Omega = -1$, so that the net Chan–Paton factor is

$$\frac{1}{2}n(n+1) - \frac{1}{2}n(n-1) = n, \quad (4.206)$$

For the $Sp(n/2)$ theory the signs are the other way around. Then the amplitude is

$$Z_{M^2} = \pm i n V_{26} \int_0^{26} \frac{dt}{4t} (8\pi^2 \alpha' t)^{-13} \vartheta_3(0, 2it)^{-12} \eta(2it)^{-12} \quad (4.207)$$

(upper sign for $SO(n)$ lower one for $Sp(n/2)$).

Seen from the closed string channel, the Möbius strip looks like a cylinder with one end closed by a cross-cap (so that we have just one boundary component). The length of the cylinder is now $\pi/4t$; by a modular transformation the amplitude becomes

$$Z_{M^2} = \pm 2 i n \frac{2^{13} V_{26}}{4\pi (8\pi^2 \alpha')^{13}} \int_0^{26} ds \vartheta_3(0, 2is/\pi)^{-12} \eta(2is/\pi)^{-12}. \quad (4.208)$$

From the asymptotics

$$\vartheta_3(0; 2is/\pi)^{-12} = 1 - 24 e^{-2s} + O(e^{-4s}) \quad (4.209)$$

$$\eta(2is/\pi)^{12} = e^{2s} + 12 e^{-2s} + O(e^{-4s}) \quad (4.210)$$

we see that the linear IR divergence is

$$\mp 2 i n \frac{24 \cdot 2^{13} V_{26}}{4\pi (8\pi^2 \alpha')^{13}} \int_0^\infty ds \quad (4.211)$$

which corresponds to a zero-momentum propagator connecting the disk tadpole with the $\mathbb{R}P^2$.

Tadpole Cancellation in Open Non-oriented Strings

For the unoriented open theory the linear divergences from the cylinder, Klein bottle, and Möbius strip sum to get

$$i \frac{24 V_{26}}{4\pi (8\pi^2 \alpha')^{13}} (2^{13} \mp n^2)^2 \int_0^\infty ds. \quad (4.212)$$

For the gauge group $SO(2^{13}) = SO(8192)$ the tadpole vanishes.

References

1. Verlinde, E., Verlinde, H.: Chiral bosonization, determinants and string partition functions. *Nucl. Phys. B* **288**, 357–396 (1987)
2. Greene, R.E., Krantz, S.G.: *Function Theory of One Complex Variable*. Graduate Studies in Mathematics, vol. 40. AMS (2006)
3. NIST Digital Library of Mathematical Functions. <https://dlmf.nist.gov>
4. Serre, J.-P.: *A Course in Arithmetics*. Springer, Berlin (1973)
5. Bruinier, J.H., van der Geer, G., Harder, G., Zagier, D.: *The 1-2-3 of Modular Forms*. Lectures at a Summer School in Nordfjordeid, Norway. Universitext, Springer (2008)
6. Diamond, F., Shurman, J.: *A First Course in Modular Forms*. Graduate Texts in Mathematics, vol. 228. Springer, Berlin (2005)
7. Gasper, G., Rahman, M.: *Basic Hypergeometric Series*. Encyclopedia of Mathematics and its Applications, vol. 96. Cambridge University Press, Cambridge (2004)
8. Alvarez-Gaume, L., Bost, J.B., Moore, G.W., Nelson, P.C., Vafa, C.: Bosonization on higher genus Riemann surfaces. *Commun. Math. Phys.* **112**, 503 (1987)
9. Alvarez-Gaumé, L., Moore, G., Vafa, C.: Theta functions, modular invariance and strings. *Commun. Math. Phys.* **106**, 1–40 (1986)
10. Schnitzer, H., Tsokos, K.: Partition functions and fermi-bose equivalence for simply-laced groups on compact Riemann surfaces. *Nucl. Phys. B* **291**, 429 (1987)
11. Rivers, R.J.: *Path Integral Methods in Quantum Field Theory*. Cambridge Monographs on Mathematical Physics, Cambridge University Press (1987)
12. Cecotti, S., Girardello, L.: Functional measure, topology and dynamical supersymmetry breaking. *Phys. Lett. B* **110**, 39 (1982)
13. Andrews, G.E., Askey, R., Roy, R.: *Special Functions*. Encyclopedia of Mathematics and its Applications, vol. 71. Cambridge University Press, Cambridge (1999)
14. Di Vecchia, P.: The Birth of String Theory. *Lectures Notes in Physics*, vol. 737, pp. 59–118 (2008). [arXiv:0704.0101](https://arxiv.org/abs/0704.0101)
15. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](https://arxiv.org/abs/1711.00864)
16. Palti, E.: The swampland: introduction and review. [arXiv:1903.06239](https://arxiv.org/abs/1903.06239)
17. Coleman, S., Weinberg, E.: Radiative corrections as the origin of spontaneous symmetry breaking. *Phys. Rev. D* **7**, 1888 (1973)
18. Cardy, J.L.: Boundary conditions, fusion rules and the Verlinde formula. *Nucl. Phys. B* **324**, 581 (1989)
19. Ishibashi, N.: The boundary and crosscap states in conformal field theories. *Mod. Phys. Lett. A* **4**, 251 (1989)
20. Bianchi, M., Sagnotti, A.: On the systematics of open string theories. *Phys. Lett. B* **247**, 517–524 (1990)
21. Angelantonj, C., Sagnotti, A.: Open strings. *Phys. Rept. C* **371**, 1–150 (2002). [arXiv:hep-th/0204089](https://arxiv.org/abs/hep-th/0204089)
22. Blumenhagen, R., Lüstring, D., Theisen, S.: *Basic Concepts of String Theory*. Springer, Berlin (2013)

Chapter 5

10d Superstring Theories



Abstract In this chapter, we construct and classify all superstring theories whose world-sheet theories have a $(1, 1)$ superconformal algebra of gauge constraints. We discuss in detail the consistency conditions of the perturbative superstring theory, namely invariance under the mapping class group of the world-sheet and absence of dangerous tadpoles/divergences. To put things in a broader perspective, we start with a review of the mapping class group. Then we show that one-loop modular invariance implies an absence of global $\mathbf{Diff}^+$ anomalies to all loop orders. We study the conditions of modular invariance in full detail. Then we compute explicitly the path integrals for Weyl fermions on a torus T^2 coupled to arbitrary flat line bundles. Finally, we work out the one-loop vacuum amplitudes for all 10d superstring models and check that they have indeed the required properties.

5.1 2d Global Gravitational Anomalies

The world-sheet QFT of a consistent (super)string theory should be $\mathbf{Diff}^+$ and **Weyl** invariant, i.e. all 2d gravitational ($\equiv \mathbf{Diff}^+$) and **Weyl** anomalies should cancel. There are two kinds of $\mathbf{Diff}^+$ anomalies: *local* and *global*. The 2d QFT is free of local gravitational anomalies iff it is invariant under “small” diffeomorphisms homotopic to the identity; in a 2d CFT this requires the left and right Virasoro central charges to be equal $c = \tilde{c}$, see **Claim 2.1**. Their common value should be *zero* if the CFT has to be Weyl invariant; this guarantees nilpotency of the BRST charge and fixes the critical dimension (cf. Chap. 3). Even when these two conditions are satisfied, the theory may fail to be invariant under homotopically *non-trivial* diffeomorphisms; we say that the world-sheet theory is plagued by *global* $\mathbf{Diff}^+$ anomalies. It follows that the local BRST conditions considered in Chap. 3 are *not* sufficient to guarantee the quantum consistency of a perturbative (super)string theory. In addition, we need two global conditions: (i) no global $\mathbf{Diff}^+$ anomalies, and (ii) absence of tadpoles that cannot be shifted away.¹

¹ The presence of a non-zero tadpole means that the “vacuum” around which we expand is not a solution to the quantum corrected e.o.m. “Shifted away” means that we can deform a little bit

There is only a handful of consistent superstring models moving in flat Minkowski space; we wish to determine their complete list by solving the global consistency conditions. In this chapter, “*superstring*” means a model whose world-sheet theory has a gauged $(1, 1)$ SUSY.² The conformal-gauge world-sheet Lagrangian is then

$$L = \frac{1}{2\pi} \left(\partial X^\mu \bar{\partial} X_\mu + \psi^\mu \bar{\partial} \psi_\mu + \tilde{\psi}^\mu \partial \tilde{\psi}_\mu + \text{ghosts} \right). \quad (5.1)$$

More general constructions of consistent string theories will be described in Chap. 7.

The Mapping Class Group of a Riemann Surface

Without loss of generality³ we assume the world-sheet Σ to be connected, oriented, and equipped with a conformal structure, hence a Riemann surface possibly with boundaries and/or punctures. The *mapping class group* $\text{MCG}(\Sigma)$ of the Riemann surface Σ is the quotient of all orientation-preserving diffeomorphisms $\Sigma \rightarrow \Sigma$ by the group of “small” diffeomorphisms isotopic to the identity, that is, [1]

$$\text{MCG}(\Sigma) \stackrel{\text{def}}{=} \pi_0(\text{Diff}(\Sigma)^+). \quad (5.2)$$

When Σ has a non-empty boundary, $\partial\Sigma \neq \emptyset$, the diffeomorphisms in $\text{Diff}(\Sigma)^+$ are required to fix the boundary *pointwise*. In the presence of punctures, the diffeomorphisms fix the punctures *as a set*, i.e. they are allowed to permute them. In the presence of p punctures, we have the exact sequence of groups

$$1 \rightarrow \text{PMCG}(\Sigma) \rightarrow \text{MCG}(\Sigma) \xrightarrow{\sigma} \mathfrak{S}_p \rightarrow 1, \quad (5.3)$$

where the $\ker \sigma \equiv \text{PMCG}(\Sigma)$ is the *pure mapping class group* of Σ .

A CFT with $c = \tilde{c}$ is free from global gravitational anomalies iff its quantum amplitudes are invariant under $\text{MCG}(\Sigma)$ for all Σ . The mapping class group of the sphere is trivial, so at the string tree level, we have no new global consistency condition.⁴ The first non-trivial condition arises at one-loop, i.e. for the torus amplitude.

Markings: A- and B-Cycles Let Σ be a closed Riemann surface of genus g . Its first homology group $H_1(\Sigma, \mathbb{Z})$ is a free Abelian group of rank $2g$: $H_1(\Sigma, \mathbb{Z}) \simeq \mathbb{Z}^{2g}$. $H_1(\Sigma, \mathbb{Z})$ is endowed with a skew-symmetric bilinear *intersection pairing*,

$$\langle \cdot, \cdot \rangle: H_1(\Sigma, \mathbb{Z}) \times H_1(\Sigma, \mathbb{Z}) \rightarrow \mathbb{Z}, \quad \langle \alpha, \beta \rangle = -\langle \beta, \alpha \rangle, \quad (5.4)$$

the vacuum configuration and get a valid solution to the required perturbative order. If this is not possible, the theory cannot be consistently defined.

² That is, locally on the world-sheet their algebra of world-sheet gauge constraints is the $(1, 1)$ SCFT algebra generated by the chiral currents $T_B(z)$, $T_F(z)$, $\tilde{T}_B(\bar{z})$, $\tilde{T}_F(\bar{z})$, see Chap. 2.

³ We can always reduce to this case by going to the oriented double.

⁴ There are no tadpoles on the sphere by Eq. (2.180).

which is unimodular⁵ by Poincaré duality [2]; we express these facts by saying that $H_1(\Sigma, \mathbb{Z})$ is a *principal symplectic lattice*. A *marking* of Σ is a choice of symplectic generators $\{A_i, B^j\}$ ($i, j = 1, \dots, g$) of the lattice $H_1(\Sigma, \mathbb{Z})$, that is, $A_i, B^j \in H_1(\Sigma, \mathbb{Z})$ such that

$$\langle A_i, A_j \rangle = \langle B^i, B^j \rangle = 0, \quad \langle A_i, B^j \rangle = -\langle B^j, A_i \rangle = \delta_i^j. \quad (5.5)$$

The A_i (resp. B^j) are called *A-cycles* (resp. *B-cycles*). Two distinct markings of Σ are related by an element of the automorphism group $Sp(2g, \mathbb{Z})$ of the symplectic lattice $H_1(\Sigma, \mathbb{Z})$. $Sp(2g, \mathbb{Z})$ is called the *Siegel modular group*.

Periods By definition, on a genus g surface Σ we have g linearly independent holomorphic differentials. The basis of such differentials, $\{\omega_i\}$ is usually chosen to be normalized with respect to the *A-cycles* so that

$$\oint_{A_i} \omega_j = \delta_j^i \quad \text{while} \quad \oint_{B^i} \omega_j = \tau_{ij} \quad (5.6)$$

The complex numbers τ_{ij} are the *periods* of Σ (with respect to the chosen marking).

Theorem 5.1 (Riemann bilinear relations [3–5]) τ_{ij} is a symmetric matrix with positive-definite imaginary part.

The space $\mathcal{H}_g \equiv \{\tau \in \mathbb{C}(g), \tau^t = \tau, \text{Im } \tau > 0\}$ of such complex $g \times g$ matrices τ is called the *Siegel upper half-space*.

Exercise 5.1 Show:⁶ (1) $Sp(2g, \mathbb{R})$ acts on $\mathcal{H}_g$ by $\tau \mapsto (A\tau + B)(C\tau + D)^{-1}$ where⁷ $\begin{pmatrix} A & B \\ C & D \end{pmatrix} \in Sp(2g, \mathbb{R})$. (2) the action is transitive. (3) the subgroup fixing $\tau \equiv i$ is $U(g)$.

The **Exercise** implies that $\mathcal{H}_g$ is (isomorphic to) the Riemannian symmetric space $Sp(2g, \mathbb{R})/U(g)$ [7]. Under a change of marking by $\gamma \in Sp(2g, \mathbb{Z}) \subset Sp(2g, \mathbb{R})$ the period matrix τ transforms according to the action in **Exercise 5.1(1)**.

Theorem 5.2 (Torelli [3–5, 8]) *The period matrix τ determines the genus g surface Σ up to isomorphism.*

For $g \geq 2$, $\dim_{\mathbb{C}} \mathcal{M}_g = 3g - 3$ while $\dim_{\mathbb{C}} \mathcal{H}_g = g(g+1)/2$, so that for $g \geq 3$ the periods of genus g surfaces form a *proper* subvariety $\mathcal{S} \subset \mathcal{H}_g$ called the *Schottky locus*. $\mathcal{S}$ is determined by a set of equations called the *Schottky relations* [8]. These complicate relations have a natural physical interpretation, see [9, 10].

⁵ An integral bilinear form $\langle \cdot, \cdot \rangle$ on a lattice Λ is *unimodular* iff it induces an isomorphism (of free Abelian groups) $\Lambda^\vee \simeq \Lambda$. Let $\{e_i\}$ be a set of generators of Λ ; the bilinear form $\langle \cdot, \cdot \rangle$ is unimodular iff $\det(e_i, e_j) = \pm 1$.

⁶ Lazy readers are referred to Sect. 1.4.2 of [6].

⁷ We use the block-matrix notation; each entry of the 2×2 “matrix” is a $g \times g$ matrix.

Symplectic Representation of the Mapping Class Group

The group $\text{MCG}(\Sigma)$ acts on the first homology group⁸ $H_1(\Sigma, \mathbb{Z})$ while preserving its intersection pairing $\langle \cdot, \cdot \rangle$. This action yields a group homomorphism from $\text{MCG}(\Sigma)$ to the automorphism group of the symplectic lattice $H_1(\Sigma, \mathbb{Z})$

$$h_*: \text{MCG}(\Sigma) \rightarrow Sp(2g, \mathbb{Z}), \quad (5.7)$$

called the *symplectic representation* of the mapping class group. The subgroup $\ker h_* \subset \text{MCG}(\Sigma)$ is called the *Torelli group* [1]. The map h_* is surjective for $g \geq 1$. When restricted to *finite* subgroups it is also injective [1].⁹ For $g = 1$, h_* is an isomorphism [1]

$$\text{MCG}(T^2) = Sp(2, \mathbb{Z}). \quad (5.8)$$

The group $Sp(2, \mathbb{Z}) \equiv SL(2, \mathbb{Z})$ is called the *modular group*. We shall show later in this section that if there are no global gravitational anomalies at one-loop ($g = 1$), there are no global anomalies to *all* loops (any g). *Absence of global Diff⁺ anomalies for all Σ is equivalent to the invariance of torus amplitudes under the modular group $SL(2, \mathbb{Z})$.* The conditions for modular invariance are a central topic in this chapter.

The Modular Group $SL(2, \mathbb{Z})$

We recall some basic facts from group theory. $SL(2, \mathbb{Z})$ is the group of integral 2×2 matrices of determinant 1. The modular group has two well-known presentations

$$SL(2, \mathbb{Z}) \equiv \langle T, L \mid TLT = LTL, (LT)^6 = 1 \rangle \quad (5.9)$$

$$\equiv \langle T, S \mid S^4 = (TS)^3 = 1, S^2T = TS^2 \rangle \quad (5.10)$$

From (5.9) it follows that the element $(LT)^3 \equiv (TL)^3$ is central; indeed

$$(LT)^3 L = L(TL)^3 \equiv L(LT)^3, \quad (LT)^3 T = (TL)^3 T = T(LT)^3, \quad (5.11)$$

and also an involution $((LT)^3)^2 = 1$. In facts $(LT)^3$ generates the center $Z \simeq \{\pm 1\}$ of $SL(2, \mathbb{Z})$, so that $PSL(2, \mathbb{Z}) \equiv SL(2, \mathbb{Z})/Z$ has the presentation

$$PSL(2, \mathbb{Z}) = \langle T, L \mid TLT = LTL, (LT)^3 = 1 \rangle. \quad (5.12)$$

Setting $S = LTL \equiv TLT$ we get $S^2 = (LT)^3$ and $TS = (TL)^2$ so that

$$(TS)^3 = (TL)^6 = (LT)^6 = 1 \quad (5.13)$$

⁸ Homology is a homotopy invariant so, if $f: \Sigma \rightarrow \Sigma$ is a diffeomorphism and α a representative of a class $[\alpha] \in H_1(\Sigma, \mathbb{Z})$, the class $[f_*\alpha] \in H_1(\Sigma, \mathbb{Z})$ depends on f only through its homotopy class.

⁹ In view of the Minkowski theorem, this implies a strong result for the mapping class group:

Theorem (Minkowski [11, 12]). *Let $\Gamma \subset GL(n, \mathbb{Z})$ be a finite subgroup. For $m \geq 3$ an integer, let $r_m: GL(n, \mathbb{Z}) \rightarrow GL(n, \mathbb{Z}/m\mathbb{Z})$ be the reduction mod m map. r_m is injective.*

Corollary. *$\text{MCG}(\Sigma)$ contains a finite-index, torsion-less, normal subgroup.*

and $L = S^{-1}TS$. In terms of explicit 2×2 matrices

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad L = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad (5.14)$$

Finally, we have¹⁰

$$PSL(2, \mathbb{Z}) = \langle S, T \mid S^2 = (TS)^3 = 1 \rangle \simeq \mathbb{Z}_2 * \mathbb{Z}_3. \quad (5.15)$$

The Artin braid group in 3 strands (i.e. the *Artin group* associated to the A_2 Dynkin graph—see **BOX 5.1**) has the presentation

$$\mathcal{B}_3 = \langle T, L \mid TLT = LTL \rangle \quad (5.16)$$

and center $Z(\mathcal{B}_3) = (TL)^{3\mathbb{Z}}$ so that

$$PSL(2, \mathbb{Z}) = \mathcal{B}_3 / Z(\mathcal{B}_3). \quad (5.17)$$

For later reference, we note the following

Lemma 5.1 *The normal closure¹¹ of T in $SL(2, \mathbb{Z})$ is the full modular group $SL(2, \mathbb{Z})$.*

Proof T and $L \equiv S^{-1}TS$ belong to the normal closure of T and generate $SL(2, \mathbb{Z})$.

Moduli of 2-Tori

As discussed in **BOX 2.4** the moduli space of complex structures on the 2-torus is

$$\mathcal{M}_{1,0} = PSL(2, \mathbb{Z}) \backslash \mathcal{H} = SL(2, \mathbb{Z}) \backslash SL(2, \mathbb{R}) / U(1) \quad (5.18)$$

where $\mathcal{H} \cong SL(2, \mathbb{R}) / U(1)$ is the upper half-plane (cf. **BOX 4.5**). $PSL(2, \mathbb{Z})$ acts on the upper half-plane by Möbius transformations

$$\tau \mapsto \frac{a\tau + b}{c\tau + d}, \quad \tau \in \mathcal{H}, \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}), \quad (5.19)$$

so the action is generated by

$$T : \tau \mapsto \tau + 1, \quad S : \tau \mapsto -1/\tau. \quad (5.20)$$

To show the absence of global **Diff**⁺ anomalies, it suffices to check invariance under T and S . A fundamental domain¹² of $PSL(2, \mathbb{Z})$ acting on the upper half-plane is

¹⁰ $*$ stands for the free *non-Abelian* product of groups.

¹¹ I.e. the smallest normal subgroup of $SL(2, \mathbb{Z})$ containing T .

¹² Cf. **Definition 4.1**.

$$F_0 = \left\{ \tau \in \mathcal{H}: -\frac{1}{2} \leq \operatorname{Re} \tau \leq \frac{1}{2}, |\tau| \geq 1 \right\} \quad (5.21)$$

see Fig. 4.3. The standard normalization of the holomorphic differential $dz/2\pi$ is such that its periods along the A - and B -cycle are

BOX 5.1 - Artin groups, Weyl groups, and Dynkin graphs

An *Artin* (or Artin-Tate) *group* A is a group with finitely many generators s_i ($i \in I$) and one relation for each pair of distinct elements $i \neq j$ of I of the form

$$\underbrace{s_i s_j s_i s_j \cdots}_{m_{i,j} \text{ factors}} = \underbrace{s_j s_i s_j s_i \cdots}_{m_{i,j} \text{ factors}}$$

where $m_{i,j} = m_{j,i} \geq 2$ are integers. An Artin group is encoded in a *diagram* Γ : Γ has a node $\bullet_i$ per each generator s_i and nodes $\bullet_i, \bullet_j$ are connected by a link with the number $m_{i,j}$

$$\cdots \cdots \bullet_i \xrightarrow{m_{i,j}} \bullet_j \cdots \cdots$$

It is conventional to delete the links with $m_{i,j} = 2$, and use special notations for $m_{i,j} = 3, 4, 6$:

$$\bullet \xrightarrow{3} \bullet \rightsquigarrow \bullet \text{---} \bullet \quad \bullet \xrightarrow{4} \bullet \rightsquigarrow \bullet \text{=} \bullet \quad \bullet \xrightarrow{6} \bullet \rightsquigarrow \bullet \text{=} \bullet$$

The *Coxeter group* W associated to A is the quotient of A by the group generated by the squares s_i^2 , i.e. the group with presentation

$$W = \left\langle s_i \ i \in I \ \middle| \ \underbrace{s_i s_j s_i s_j \cdots}_{m_{i,j} \text{ factors}} = \underbrace{s_j s_i s_j s_i \cdots}_{m_{i,j} \text{ factors}} \text{ and } s_i^2 = 1, \forall i, j \in I \right\rangle.$$

When Γ is the Dynkin graph of a Lie algebra, W is its *Weyl group*.

$$\oint_A \frac{dz}{2\pi} = 1, \quad \oint_B \frac{dz}{2\pi} = \tau, \quad (5.22)$$

so the modular action on τ (5.19) coincides with the natural symplectic action of $SL(2, \mathbb{Z})$ on $H_1(\Sigma, \mathbb{Z}) = \mathbb{Z}A \oplus \mathbb{Z}B$.

More on the Mapping Class Group

Consider the finite cylinder $\text{Cy} = [0, 1] \times S^1$ with coordinates (x, θ) . $\text{MCG}(\text{Cy})$ is isomorphic to $\mathbb{Z}$ with generator the *Dehn twist*, i.e. the diffeomorphism [1]

$$(x, \theta) \mapsto (x, \theta + 2\pi x), \quad (5.23)$$

which is topologically non-trivial and fixes the boundary ∂Cy pointwise.

Let Σ be an oriented surface, possibly with punctures and boundaries, and $\alpha \subset \Sigma$ a simple closed curve. Let $N_\alpha \subset \Sigma$ be a tubular neighborhood of α . N_α is diffeomorphic to $\mathbf{C}y$. The *Dehn twist along the curve α* is the diffeomorphism $\Sigma \rightarrow \Sigma$ which is the Dehn twist in $N_\alpha \simeq \mathbf{C}y$ and the identity on the complement $\Sigma \setminus N_\alpha$. We write $T_\alpha \in \mathbf{MCG}(\Sigma)$ for the class of the Dehn twist along α . Isotopic curves define the same element $T_\alpha \in \mathbf{MCG}(\Sigma)$. T_α is non-trivial unless α is homotopic to a point or to a puncture [1]. Dehn twists T_α, T_β along two non-intersecting curves α, β commute. If they intersect exactly once, they satisfy the braid relation [1]

$$T_\alpha T_\beta T_\alpha = T_\beta T_\alpha T_\beta. \quad (5.24)$$

For instance on the torus T^2 the Dehn twists along the simple closed curves in the homology classes $(1, 0)$ and $(0, 1)$ are the elements T^{-1} and L^{-1} of $Sp(2, \mathbb{Z}) \simeq \mathbf{MCG}(T^2)$; cf. Eq. (5.9). Finally, the Dehn twists along any two non-separating¹³ simple closed curves are conjugate in $\mathbf{MCG}(\Sigma)$; cf. Sect. 1.3.1 of [1].

Theorem 5.3 (Dehn-Lickorish—Theorem 4.11 in [1]) *Let Σ be a closed oriented surface of genus $g \geq 1$ and $n \geq 0$ punctures. Then the group $\mathbf{PMCG}(\Sigma)$ is generated by finitely-many Dehn twists along non-separating simple closed curves. The same holds in the presence of boundaries for $g \geq 2$ while for $g = 1$ we need also the Dehn twists along $b - 1$ separating curves where b is the number of boundary components.*

Modular Invariance at Higher Genus

We justify our claim that modular invariance on the torus implies an absence of global gravitational anomalies in all world-sheet topologies. We may assume Σ to be connected, closed, and orientable of genus $g \geq 2$. Topologically Σ is a connected sum of g tori (cf. BOX 1.5)

$$\Sigma = \overbrace{T^2 \# T^2 \# \dots \# T^2}^{g \text{ copies}} \quad (5.25)$$

This means that we may draw $g - 1$, pairwise non-intersecting, simple closed curves $\gamma_a \subset \Sigma$ such that cutting the surface along the γ_a 's we remain with a disconnected collection of tori $\mathcal{T}_i$ with disks removed. Let us focus on one of these tori, $\mathcal{T}_{i_0}$. We can split the path integral on Σ (with insertions) in three distinct functional integrals:

(1) the path integral on $\mathcal{T}_{i_0}$ over the fields¹⁴ satisfying the boundary conditions

$$\Phi(z) \Big|_{\partial \mathcal{T}_{i_0}} = \phi(z) \equiv \text{a fixed field configuration on } \partial \mathcal{T}_{i_0} \quad (5.26)$$

¹³ A closed curve is non-separating if cutting along it the surface Σ remains connected.

¹⁴ The symbol Φ stands for the collection of all fields we integrate over in the path integral.

- (2) the path integral over the rest of the surface, $\check{\Sigma}_{i_0} \equiv \Sigma \setminus \mathcal{T}_{i_0}$, with the same boundary condition (5.26);
- (3) the functional integral over the boundary configuration $\phi(z)$ of the fields.

The path integral on $\check{\Sigma}_{i_0}$ yields the Schrödinger representation of a state in the Hilbert space associated to its boundary $\partial\mathcal{T}_{i_0}$

$$|\Psi\rangle \equiv \sum_{\alpha} \bigotimes_{s=1}^{b_{i_0}} |\psi_{\alpha,s}\rangle \in \mathbf{H}^{\otimes b_{i_0}}, \quad (5.27)$$

$$b_{i_0} \equiv \#(\text{connected components of } \partial\mathcal{T}_{i_0}),$$

where $\mathbf{H}$ is the Hilbert space of normalizable functionals $\Psi[\Phi(\theta)]$ of the field configurations on S^1 ; cf. the proof of the state-operator isomorphism in Sect. 2.2.2. If all operators inserted in the original surface Σ are local¹⁵ and BRST-invariant, $|\Psi\rangle$ and hence each $|\psi_{\alpha,s}\rangle$ is BRST-invariant; this follows from the fact that under these assumptions the BRST currents are global meromorphic 1-forms with poles only at the operator insertions and zero residues, while $\partial\check{\Sigma}_{i_0}$ is homologous to the sum of small loops encircling the operator insertions in $\check{\Sigma}_{i_0}$, so that

$$Q_{\text{BRS}}|\Psi\rangle = \sum_{x: \text{ insertions in } \check{\Sigma}_{i_0}} (\text{residues of BRST current at } x) = 0. \quad (5.28)$$

By the state-operator correspondence, the state $|\Psi\rangle$ is also produced by a sum of path integrals over a disconnected collection of b_{i_0} disks where in the center of the s -th disk we insert the BRST-invariant operator $\psi_{\alpha,s}$. The original amplitude $A(\Sigma)$ is given by the sum $\sum_{\alpha} A(\overline{\mathcal{T}}_{i_0}; \alpha)$ of amplitudes $A(\overline{\mathcal{T}}_{i_0}; \alpha)$ computed on the closed torus $\overline{\mathcal{T}}_{i_0}$ obtained by filling in each boundary component of $\mathcal{T}_{i_0}$ with a disk and inserting the operator $\psi_{\alpha,s}$ in the center of the s -th disk (in addition to the operators already present in $\mathcal{T}_{i_0} \subset \Sigma$).

Now assume that the torus amplitudes with arbitrary BRST-invariant local insertion are modular invariant. Then $A(\overline{\mathcal{T}}_{i_0}; \alpha)$ is invariant under the Dehn twist T_{ℓ} along any closed simple curve $\ell \subset \mathcal{T}_{i_0} \subset \overline{\mathcal{T}}_{i_0}$. Replacing the filling disks with the rest of the surface $\check{\Sigma}_{i_0}$, ℓ and T_{ℓ} get identified with a closed simple curve and the associated Dehn twists on the original surface Σ . The invariance of $A(\overline{\mathcal{T}}_{i_0}; \alpha)$ under T_{ℓ} implies the invariance of the original multi-loop amplitude $A(\Sigma)$ under T_{ℓ} .

By **Theorem 5.3**, the group $\text{PMCG}(\Sigma)$ is generated by Dehn twists along non-separating closed simple curves ℓ . Each such curve is isotopic to some non-separating closed simple curve ℓ' contained in a torus $\mathcal{T}_i$ for some decomposition of the surface Σ in a collection of tori $\Sigma \simeq \#_i \mathcal{T}_i$; this follows from the fact that the mapping class group acts transitively on the isotopy classes of non-separating, closed, simple

¹⁵ By *local* we mean that each operator is local with respect to all inserted operators (including itself) and also with respect to the BRST currents (left and right).

curves. We conclude¹⁶ that if the torus amplitudes with arbitrary mutually-local BRST-invariant insertions are modular invariant then the genus $g \geq 1$ amplitudes with arbitrary mutually-local BRST-invariant insertions is invariant for $\text{MCG}(\Sigma)$.

Operator Perspective on All-Loop $\text{MCG}(\Sigma)$ Invariance

We consider a genus g surface Σ with arbitrary local BRST-invariant insertions. Let $\ell \subset \Sigma$ be a closed simple curve. The Dehn twist T_ℓ is a non-trivial element of $\text{MCG}(\Sigma)$ if and only if ℓ does not bound a disk or a once-punctured disk [1]. We know that the T_ℓ 's generate $\text{MCG}(\Sigma)$ (a finite subset of them suffices, see Theorem 5.3). We cut the surface Σ along ℓ ; we get either two disconnected surfaces Σ_1, Σ_2 of genus g_1 and g_2 (with $g_1 + g_2 = g$) and one boundary each, or a connected surface $\tilde{\Sigma}$ of genus $g - 1$ with two boundary components. In the first, resp. second case, the quantum amplitude has schematically the structure¹⁷

$$\sum_{\psi \in \mathbf{H}} \langle \Sigma_1 | \psi \rangle \langle \psi | \Sigma_2 \rangle \quad \text{resp.} \quad \sum_{\psi \in \mathbf{H}} \langle \psi | \tilde{\Sigma} | \psi \rangle, \quad (5.29)$$

where $|\Sigma_a\rangle \in \mathbf{H}$ is the state produced by the path integral over Σ_a (for an appropriate orientation of its boundary) and $\tilde{\Sigma}$ is the vector in $\mathbf{H} \otimes \mathbf{H}^\vee \simeq \mathbf{H}^{\otimes 2}$ produced by the path integral over $\tilde{\Sigma}$ with opposite orientation for its two boundaries.

By its very definition, T_ℓ acts on $\mathbf{H}$ as the operator $\exp[-2\pi i(L_0 - \tilde{L}_0)]$, so after the Dehn twist along ℓ the amplitude becomes

$$\sum_{\psi \in \mathbf{H}} \langle \Sigma_1 | e^{-2\pi i(L_0 - \tilde{L}_0)} | \psi \rangle \langle \psi | \Sigma_2 \rangle \quad \text{resp.} \quad \sum_{\psi \in \mathbf{H}} \langle \psi | \tilde{\Sigma} e^{-2\pi i(L_0 - \tilde{L}_0)} | \psi \rangle, \quad (5.30)$$

and invariance requires $\exp[2\pi i(L_0 - \tilde{L}_0)]$ to act on $\mathbf{H}$ as the identity operator, i.e.

$$L_0 - \tilde{L}_0 \in \mathbb{Z} \quad \text{for all states in } \mathbf{H}. \quad (5.31)$$

Naively Eq. (5.31) is the *only* condition for modular invariance of all physical amplitudes to all orders in the genus expansions; indeed (5.31) guarantees invariance under arbitrary Dehn twists, and they generate $\text{MCG}(\Sigma)$. However there are other subtler conditions which are hidden under our cavalier manipulations of the path integral; we have to ascertain that our *formal* treatment of the quantum amplitudes is justified.

First of all, we must preserve 2d conformal invariance to have a state-to-operator correspondence; the Hilbert space $\mathbf{H}$ should be linearly isomorphic to an algebra $\mathfrak{A}$ of mutually-local operators $\mathcal{O}(z, \bar{z})$. Locality and closure of the OPE algebra $\mathfrak{A}$

¹⁶ The argument shows invariance under the *pure* group $\text{PMCG}(\Sigma)$. To show that it is invariant under the full mapping class group $\text{MCG}(\Sigma)$, we have to show that it is also invariant under the symmetric group $\mathfrak{S}_n$ permuting the n punctures. This amounts to showing that the inserted operators have bosonic statistics, i.e. (by the 2d Spin & Statistics theorem) that their 2d spins are integral. This is an automatic consequence of BRST invariance or of left-right matching conditions.

¹⁷ As before, $\mathbf{H}$ is the physical Hilbert space of the closed string.

set strong constraints on the allowed $\mathbf{H}$'s. This is not enough; not all local operator algebras $\mathfrak{A}$ satisfying (5.31), that is, with

$$e^{-2\pi i(L_0 - \tilde{L}_0)} O(0) e^{2\pi i(L_0 - \tilde{L}_0)} = O(0) \quad \text{for all } O(z, \bar{z}) \in \mathfrak{A} \quad (5.32)$$

will do. For instance, all proper subalgebras $\mathfrak{A}' \subsetneq \mathfrak{A}$ tautologically satisfy the above condition, except that

$$\sum_{\phi \in \mathfrak{A}'} \langle \Sigma_1 | \phi \rangle \langle \phi^\dagger | \Sigma_2 \rangle \neq \sum_{\phi \in \mathfrak{A}} \langle \Sigma_1 | \phi \rangle \langle \phi^\dagger | \Sigma_2 \rangle \equiv \text{path integral on } \Sigma \text{ with insertions} \quad (5.33)$$

so that the "small" operator algebra $\mathfrak{A}'$ does not realize the path integral, and our formal arguments in Eqs. (5.29)–(5.31) would be invalid; cf. the discussion in Sect. 2.3.7. It is clear that we need to supplement the condition (5.31) with a *completeness* requirement on the operator algebra $\mathfrak{A}$, that is,

$$\sum_{\phi \in \mathfrak{A}} |\phi\rangle \langle \phi^\dagger| = \text{Id}_{\mathbf{H}}, \quad (5.34)$$

which guarantees *inter alia* that

$$\sum_{\phi \in \mathfrak{A}} \langle \phi^\dagger | q^{L_0 - c/24} \bar{q}^{\tilde{L}_0 - c/24} | \phi \rangle \equiv \text{path integral over the torus}, \quad (5.35)$$

i.e. that the operator algebra $\mathfrak{A}$ *does realize* the path integral. The completeness criterion (5.34) may be rephrased as a *maximality* condition on the algebra $\mathfrak{A}$; if there is a legitimate local operator O_* that we may possibly insert in the path integral, its functional Schrödinger representation $\langle \Phi(\theta) | O_* \rangle$ makes sense, hence so does the state $|O_*\rangle \in \mathbf{H}$. Then O_* should appear in the resolution of the identity (5.34), and hence O_* *must* belong to $\mathfrak{A}$. Thus *all* operators which are mutually local with the operators in $\mathfrak{A}$ should be in $\mathfrak{A}$: $\mathfrak{A}$ is a *maximal local operator algebra*. Conversely, in 2d CFT all maximal local operator algebras $\mathfrak{A}$ satisfy the completeness condition (5.34). Indeed, suppose it is not so, then $\ker(\text{Id}_{\mathbf{H}} - \sum_{\phi \in \mathfrak{A}} |\phi\rangle \langle \phi^\dagger|)$ is a non-zero subspace of states which corresponds to a linear space of local operators not in $\mathfrak{A}$ but mutually local with respect to $\mathfrak{A}$, contradicting maximality. We conclude:

A 2d CFT, with $c = \tilde{c}$ and local operator algebra $\mathfrak{A}$, has $\text{MCG}(\Sigma)$ -invariant amplitudes for all surfaces Σ and Diff^+ -invariant insertions if and only if:

- (i) $\mathfrak{A}$ is maximal with respect to locality
- (ii) $\exp[-2\pi i(L_0 - \tilde{L}_0)] O(0) \exp[2\pi i(L_0 - \tilde{L}_0)] = O(0)$ for all $O(z, \bar{z}) \in \mathfrak{A}$

Note 5.1 Condition (i) implies $h - \tilde{h} \in \frac{1}{2}\mathbb{Z}$ (from the locality of each operator with itself). Condition (ii) requires the slightly stronger condition $h - \tilde{h} \in \mathbb{Z}$. We give a counter-example: consider 8 left-moving plus 8 right-moving MW fermions $\psi^i(z)$, $\tilde{\psi}^i(\bar{z})$ ($i = 1, \dots, 8$), and perform independent GSO projections on the left and right sides. From Sect. 2.9, we know that the resulting OPE algebra is maximal local. But the left-moving spin fields $S_\alpha(z)$ have $(h, \tilde{h}) = (\frac{1}{2}, 0)$, violating (ii), so the torus amplitudes are *not* modular invariant. Indeed T acts on $\mathbf{H}$ as the diagonal 2π rotation $(-1)^F \in Spin(8)_L \times Spin(8)_R$. We restore modular invariance by twisting the amplitudes with some *extra signs*; for a deeper perspective see Remark 5.5. On the contrary for $16 + 16$ MW fermions, the GSO-projected algebra is maximal local with integral spins since $S_\alpha(z)$ has $(h, \tilde{h}) = (1, 0)$.

Modular Transformations of Chiral Partition Functions

Suppose we have a CFT algebra $\mathfrak{A}$ which satisfy conditions (i), (ii) but have $c \neq \tilde{c}$, so that we have *local* $\mathbf{Diff}^+$ anomalies. In this case, the torus amplitudes are *not* modular invariant. However, maximal locality and spin integrality imply simple formulae for the modular transformations of the torus partition function

$$Z_{\mathfrak{A}}(\tau + 1) = e^{2\pi i(\tilde{c}-c)/24} Z_{\mathfrak{A}}(\tau), \quad Z_{\mathfrak{A}}(-1/\tau) = Z_{\mathfrak{A}}(\tau). \quad (5.36)$$

The proof of (5.36) works as in the previous paragraph, except that the Dehn twist now acts on $\mathbf{H}$ as multiplication by $\exp[2\pi i(c - \tilde{c})/24]$. Equations (5.36) hold for all torus amplitude, not just the partition function, as long as the insertions are $\mathbf{Diff}^+$ -invariant (as the vertices in string theory). This result applies, in particular, to *chiral* CFT with only left-moving d.o.f. $c > 0$ and $\tilde{c} = 0$. We conclude

Fact 5.1 A (left-moving) chiral CFT is modular invariant iff its OPE algebra $\mathfrak{A}$ is maximal local with integral spins and $c = 24n$ for $n \in \mathbb{N}$.

The anomalous phase in the modular transformation gives a one-dimensional representation χ of the modular group; hence it factorizes through its Abelianization¹⁸

$$PSL(2, \mathbb{Z}) \rightarrow PSL(2, \mathbb{Z})^{\text{ab}} \rightarrow U(1) \quad (5.37)$$

From Eq. (5.15) we see that $PSL(2, \mathbb{Z})^{\text{ab}} = \mathbb{Z}_2 \times \mathbb{Z}_3 \simeq \mathbb{Z}_6$. Hence the image of χ is contained in the group μ_6 of 6-th roots of unity. But maximal locality implies invariance under S , so we have the stronger result $\chi \in \mu_3$. Since the image is generated by $\exp[2\pi i(\tilde{c} - c)/24]$ we conclude

Fact 5.2 If $\mathfrak{A}$ is maximal local with integral spin, then $c - \tilde{c} \in 8\mathbb{Z}$.

¹⁸ The Abelianization G^{ab} of a group G is the quotient group $G^{\text{ab}} \stackrel{\text{def}}{=} G/[G, G]$, where $[G, G]$ is the normal subgroup generated by the commutators.

5.2 Consistent Closed Superstring Theories in 10d

We consider first the closed oriented superstring theories with the world-sheet Lagrangian (5.1). As discussed in Chap. 2, in the world-sheet SCFT we have $2^4 = 16$ Hilbert space sectors labeled by

$$(\alpha, F; \tilde{\alpha}, \tilde{F}) \quad (5.38)$$

where α (resp. $\tilde{\alpha}$) distinguishes the NS sector from the R sector for left-movers (resp. right-movers)

$$\alpha \equiv 1 - 2\nu = \begin{cases} 1 & \text{R sector} \\ 0 & \text{NS sector,} \end{cases} \quad (5.39)$$

while F (resp. $\tilde{F}$) counts the left-moving (resp. right-moving) fermion number mod 2. We write $NS(-1)^F$ and $R(-1)^F$ for $(0, F)$ and $(1, F)$, respectively.

BOX 5.2 - Mutual locality of world-sheet operators

Let us illustrate Eq. (5.40) using bosonization. Ignoring the ∂X^μ , b , c oscillators which are the same in all sectors, we may write the operators in the $SO(10, 2)$ bosonized form

$$\mathcal{O}_{\lambda, \tilde{\lambda}}(z, \bar{z}) = c_{\lambda, \tilde{\lambda}} e^{i\lambda \cdot \phi(z) + i\tilde{\lambda} \cdot \tilde{\phi}(\bar{z})}$$

where $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5; \lambda_6)$ with integer entries in the NS sector and half-integer in the R one, i.e. $\lambda_i = \frac{1}{2}\alpha \bmod 1$, while $F = \sum_i \lambda_i \bmod 2$ (and the same for tilted quantities). We write

$$\lambda = (\tfrac{1}{2}\alpha, \tfrac{1}{2}\alpha, \tfrac{1}{2}\alpha, \tfrac{1}{2}\alpha, \tfrac{1}{2}\alpha; -\tfrac{1}{2}\alpha) + n, \quad n \in \mathbb{Z}^6, \quad F \equiv \sum_i n_i,$$

and

$$\lambda_1 \cdot \lambda_2 = \alpha_1 \alpha_2 + \tfrac{1}{2}(\alpha_1 F_2 + \alpha_2 F_1) + n_1 \cdot n_2 = \tfrac{1}{2}(\alpha_1 F_2 - \alpha_2 F_1) \bmod 1,$$

where $\cdot$ is the Lorentzian product in Sect. 3.1 Then

$$\mathcal{O}_{\lambda_1, \tilde{\lambda}_1}(z, \bar{z}) \mathcal{O}_{\lambda_2, \tilde{\lambda}_2}(0) = (\text{cocycl.}) z^{\lambda_1 \cdot \lambda_2} \bar{z}^{\tilde{\lambda}_1 \cdot \tilde{\lambda}_2} e^{i(\lambda_1 \cdot \phi(z) + \lambda_2 \cdot \phi(0) + \tilde{\lambda}_1 \cdot \tilde{\phi}(\bar{z}) + \tilde{\lambda}_2 \cdot \tilde{\phi}(0))}$$

Setting $z = e^{2\pi i}$, the overall phase as the first operator is circled around the second is

$$\exp[2\pi i(\lambda_1 \cdot \lambda_2 - \tilde{\lambda}_1 \cdot \tilde{\lambda}_2)] = \exp[\pi i(\alpha_1 F_2 - \alpha_2 F_1 - \tilde{\alpha}_1 \tilde{F}_2 + \tilde{\alpha}_2 \tilde{F}_1)]$$

To get a consistent superstring theory we must combine the several sectors into a Hilbert space $\mathbf{H}$ consistent with the fundamental principles of string theory. By the CFT state-operator isomorphism, a consistent superstring Hilbert space is best seen as a consistent algebra of local operators $\mathfrak{A}$, its *vertex algebra*. $\mathfrak{A}$ is a subalgebra of the BRST cohomology algebra $H_{\text{BRS}}(\mathcal{V}^{\text{oper}})$, closed under Hermitian conjugation and picture changing, which carries a representation of the 10d Poincaré group.

Each physical vertex $O_{(\alpha, F; \tilde{\alpha}, \tilde{F})} \in H_{\text{BRS}}(\mathcal{V}^{\text{oper}})$ comes with labels $(\alpha, F; \tilde{\alpha}, \tilde{F})$ which specify its sector. The labels $(\alpha, F; \tilde{\alpha}, \tilde{F})$ are additive mod 2 in OPEs. To lead to unambiguous amplitudes, the algebra $\mathfrak{A}$ must be *local*, i.e. any two operators in $\mathfrak{A}$ must be mutually local. In the presence of an R sector operator, the fermions have square-root branch-cuts. Since F counts fermions mod 2, an operator crossing the branch-cut picks up the phase $e^{\pm i\pi F}$. The net phase acquired when a $(\alpha_2, F_2; \tilde{\alpha}_2, \tilde{F}_2)$ operator goes around a $(\alpha_1, F_1; \tilde{\alpha}_1, \tilde{F}_1)$ operator is then

$$\exp\left[\pi i(F_1\alpha_2 - F_2\alpha_1 - \tilde{F}_1\tilde{\alpha}_2 + \tilde{F}_2\tilde{\alpha}_1)\right] \quad (5.40)$$

cf. **BOX 5.2** for a check using bosonized vertices. If the two sectors $(\alpha_1, F_1; \tilde{\alpha}_1, \tilde{F}_1)$ and $(\alpha_2, F_2; \tilde{\alpha}_2, \tilde{F}_2)$ are both in $\mathbf{H}$, we must have

$$\alpha_1 F_2 - \alpha_2 F_1 - \tilde{\alpha}_1 \tilde{F}_2 + \tilde{\alpha}_2 \tilde{F}_1 \in 2\mathbb{Z}. \quad (5.41)$$

Finally, $\mathfrak{A}$ should be consistent with modular invariance, i.e. $\mathfrak{A}$ must be a maximal local operator algebra with $h - \tilde{h} \in \mathbb{Z}$. This leads to three *necessary* conditions:

- (1) the sector $(NS+, NS+)$ must be in $\mathbf{H}$. Indeed $1 \in \mathfrak{A}$ and $1 \in (NS+, NS+)$;
- (2) $NS-$ pairs only with itself: $\tilde{h} \in \frac{1}{2} + \mathbb{Z}$ in $NS-$, while $h \in \mathbb{Z}$ in $NS+, R\pm$;
- (3) the left- or right-moving sector cannot consist only of $NS\pm$ sectors; otherwise the field $\psi^\mu(z)$ with $h - \tilde{h} = \frac{1}{2} \notin \mathbb{Z}$ is local with all operator, hence in $\mathfrak{A}$.

We now list the operator algebras $\mathfrak{A}$ which enjoy these conditions.

BOX 5.3 - *Spin & Statistics theorem in physical 10d spacetime*

In the bosonized formulation, an operator O of momentum p , pictures $(q, \tilde{q})$ and ghost extended weights $(\lambda, \tilde{\lambda})$ (with $\lambda = (\lambda_1, \dots, \lambda_5; q)$; cf. Sect. 3.1) has

$$h = \frac{1}{2}p^2 + \frac{1}{2}\lambda \cdot \lambda - q + N, \quad \tilde{h} = \frac{1}{2}p^2 + \frac{1}{2}\tilde{\lambda} \cdot \tilde{\lambda} - \tilde{q} + \tilde{N},$$

where $N, \tilde{N}$ are the left/right oscillator levels. The level-matching condition $h = \tilde{h}$ implies

$$\lambda \cdot \lambda - \tilde{\lambda} \cdot \tilde{\lambda} = 2(q - \tilde{q}) \pmod{2} \quad (\star)$$

From the form of the vertices, we see that under a spacetime rotation by 2π a physical state picks up the phase $e^{2\pi i(q - \tilde{q})}$ so it has spin $(q - \tilde{q}) \pmod{1}$. On the other hand,

$$O(z, \bar{z})O(w, \bar{w}) = (z - w)^{\lambda \cdot \lambda}(\bar{z} - \bar{w})^{\tilde{\lambda} \cdot \tilde{\lambda}}(\dots) = (-1)^{\lambda \cdot \lambda - \tilde{\lambda} \cdot \tilde{\lambda}}O(w, \bar{w})O(z, \bar{z})$$

so a state is bosonic (fermionic) if $\lambda \cdot \lambda - \tilde{\lambda} \cdot \tilde{\lambda}$ is 0 (resp. 1) mod 2. Then Eq. $(\star)$ expresses the *Spin & Statistics theorem* in spacetime.

Closed Superstrings with Spacetime Fermions

We consider first the case where the 10d spectrum contains spacetime spinors belonging, say, to a R-NS sector $(\alpha, \tilde{\alpha}) = (1, 0)$. By condition (2) the right part of this sector should be NS+, so this sector is either (R+,NS+) or (R-,NS+), and only one of the two possibilities appears since the two are not mutual local (cf. Sect. 3.1). By condition (3) there must also be a NS-R or a R-R sector; since $R-R \times R-NS = NS-R$ in both cases, there is a NS-R sector. Again, this sector should be (NS+,R+) or (NS+,R-), but not both. Thus we have four possibilities for spacetime fermions: $(R_{\varepsilon_L}, NS+)$ and $(NS+, R_{\varepsilon_R})$, where $\varepsilon_L, \varepsilon_R = \pm$. Closure of the algebra then requires also the sectors

$$(R_{\varepsilon_L}, NS+) \times (R_{\varepsilon_L}, NS+) = (NS+, NS+) \quad (5.42)$$

$$(R_{\varepsilon_L}, NS+) \times (NS+, R_{\varepsilon_R}) = (R_{\varepsilon_L}, R_{\varepsilon_R}) \quad (5.43)$$

Let $\mathfrak{A}_{\varepsilon_L, \varepsilon_R}$ be the algebra of all BRST-invariant operators in the four sectors

$$(NS+, NS+), (R_{\varepsilon_L}, NS+), (NS+, R_{\varepsilon_R}), (R_{\varepsilon_L}, R_{\varepsilon_R}), \quad (5.44)$$

for a choice of the signs $\varepsilon_L, \varepsilon_R$. $\mathfrak{A}_{\varepsilon_L, \varepsilon_R}$ is *maximally local*, i.e. if a BRST-invariant operator $\mathcal{O}$ is local with respect to all operators in $\mathfrak{A}_{\varepsilon_L, \varepsilon_R}$ then $\mathcal{O} \in \mathfrak{A}_{\varepsilon_L, \varepsilon_R}$. We conclude: *if the modular-invariant vertex algebra $\mathfrak{A}$ contains a spacetime fermion¹⁹ vertex, it should be one of the four $\mathfrak{A}_{\varepsilon_L, \varepsilon_R}$.*

We read the physical spectrum of each consistent model directly from its vertex algebra $\mathfrak{A}_{\varepsilon_L, \varepsilon_R}$. We stress that none of the four spectra (5.44) contains the closed string tachyon which belongs to the $(NS-, NS-)$ sector; the tachyon vertex is $e^{-\phi - \tilde{\phi}} e^{ik \cdot X}$, so it has left and right Fermi numbers -1 ; cf. Sect. 3.1.

By inspection, we see that the four consistent algebras $\mathfrak{A}_{\varepsilon_L, \varepsilon_R}$ which contain fermion vertices are obtained by *independent* GSO $\pm$ projections on the left- and the right movers, i.e. by the projection

$$P_{\varepsilon_L, \varepsilon_R} \equiv \frac{1}{2} (1 + \varepsilon_L (-1)^F) \cdot \frac{1}{2} (1 + \varepsilon_R (-1)^{\tilde{F}}). \quad (5.45)$$

These four algebras represent just *two physically distinct theories*; indeed, a space-time reflection on a single axis

$$X^9 \rightarrow -X^9, \quad \psi^9 \rightarrow -\psi^9, \quad \tilde{\psi}^9 \rightarrow -\tilde{\psi}^9 \quad (5.46)$$

leaves the action and the constraints unchanged but interchanges $R\pm \leftrightarrow R\mp$ simultaneously for the left- and the right-movers. What matters is whether the independent GSO projections on the two sides preserve *opposite* chiralities $\varepsilon_L = -\varepsilon_R$ or the *same* one $\varepsilon_L = \varepsilon_R$. The model with opposite projections is called *Type IIA superstring*, and the one with equal projections *Type IIB superstring*. Their sector content is:

¹⁹ We use “fermion” and “spinor” interchangeably since one consequence of the analysis is that the Spin & Statistic Theorem holds in 10d, see BOX 5.3.

Model	Sectors				(5.47)
IIA	(NS+,NS+)	(R+,NS+)	(NS+,R-)	(R+,R-)	
IIB	(NS+,NS+)	(R+,NS+)	(NS+,R+)	(R+,R+)	

Exercise 5.2 Using the bosonized formulation, show that spacetime parity interchanges $R_{\pm} \leftrightarrow R_{\mp}$ on both sides of the closed superstring.

Closed Superstrings Without Spacetime Fermions

Next we assume that no NS-R nor R-NS sector is present. By condition (1), $(NS+, NS+)$ is present. By condition (3), there is an R-R sector, say $(R_{\varepsilon_L}, R_{\varepsilon_R})$. Mutual locality then forbids the sectors $(R - \varepsilon_L, R_{\varepsilon_R})$ and $(R_{\varepsilon_L}, R - \varepsilon_R)$ but allows the sector $(R - \varepsilon_L, R - \varepsilon_R)$. By maximality of $\mathfrak{A}$, we have to add this sector. Closure of the OPE then requires $(NS-, NS-)$. The resulting algebra is maximal. Again we have two physically inequivalent models depending on the relative sign $\varepsilon_R/\varepsilon_L$:

Model	Sectors				(5.48)
0A	(NS+,NS+)	(NS-,NS-)	(R+,R-)	(R-,R+)	
0B	(NS+,NS+)	(NS-,NS-)	(R+,R+)	(R-,R-)	

They are both affected by the presence of a tachyon.

Discussion: Physical Spectrum, SUSY, and All That

We have found two interesting closed superstring theories, *Type IIA* and *Type IIB*, with fermions in their physical spectrum and no tachyon. These models are fully consistent in string perturbation theory. The physical massless spectra of these two nice theories, written in terms of representations of the Lorentz small group $Spin(8)$, may be read from Table 3.1 which we reproduce for the convenience of the reader:

Model	bosons	fermions	(5.49)
<i>IIA</i>	$\mathbf{Sy} \oplus \Lambda_0 \oplus \Lambda_1 \oplus \Lambda_2 \oplus \Lambda_3$	$\mathbf{8}_s \oplus \mathbf{8}_c \oplus \mathbf{56}_s \oplus \mathbf{56}_c$	
<i>IIB</i>	$\mathbf{Sy} \oplus \Lambda_0^{\oplus 2} \oplus \Lambda_2^{\oplus 2} \oplus \Lambda_4^+$	$\mathbf{8}_c^{\oplus 2} \oplus \mathbf{56}_s^{\oplus 2}$	

Type IIB is obtained by the GSO projections which keep all sectors with

$$\Gamma = \tilde{\Gamma} = +1, \quad (5.50)$$

while Type IIA by the GSO projections which keep

$$\Gamma = +1 \quad \tilde{\Gamma} = -1. \quad (5.51)$$

The spacetime spectrum of Type IIA in *non-chiral*, i.e. invariant under spacetime parity which interchanges $\mathbf{8}_s \leftrightarrow \mathbf{8}_c$ and $\mathbf{56}_s \leftrightarrow \mathbf{56}_c$. On the world-sheet, this symmetry is the product of the spacetime parity (5.46) and the world-sheet parity Ω . Type IIB has instead a *chiral* spectrum; chiral fermions but also *chiral bosons*, in the sense

of a 4-form gauge field $A^{(4)}$ whose field strength 5-form $F^{(5)} = dA^{(4)}$ is constrained to be self-dual.²⁰

On the contrary, Type 0 is obtained by projections that correlate the left- and the right movers to be in the same sector (0B) or in the parity-reversed sector (0A).

Supersymmetry The most important property of Type II theories is that their R-NS and NS-R sectors contain massless spin- $\frac{3}{2}$ gravitini with vertices

$$S_\alpha \tilde{\psi}_\mu e^{-\phi/2-\tilde{\phi}} e^{ik \cdot X}, \quad \tilde{S}_\alpha \psi_\mu e^{-\tilde{\phi}/2-\phi} e^{ik \cdot X}. \quad (5.52)$$

We proved in Sect. 3.8 that models with independent left-/right- GSO projections are automatically *supersymmetric*, while SUSY is a consistency requirement in the presence of massless gravitini [13]. The absence of tachyons is also a consequence of SUSY. The name *Type II* (resp. *Type 0*) counts the number of supercharges in spacetime, which are 2 (resp. 0) 10d Majorana-Weyl (MW) spinors, corresponding to the two gravitini, one from the R-NS sector and one from the NS-R one. More precisely:

- in Type IIA we have a supercharge transforming in the $\mathbf{16}_s$ of $SO(9, 1)$ (i.e. a positive chirality 10d MW spinor) and one in the $\mathbf{16}_c$ (i.e. a negative chirality 10d MW spinor);
- in Type IIB both supercharges are in the $\mathbf{16}_s$ of $SO(9, 1)$, i.e. they are two 10d MW spinors of the *same* chirality.

5.3 Consistent Unoriented and Open Superstrings

Closed Unoriented Superstrings

Type IIB superstring has the same GSO projection on both sides and hence has a world-sheet parity symmetry Ω interchanging left- and right movers. We can gauge this $\mathbb{Z}_2$ symmetry to get an *unoriented* closed superstring theory. In the massless NS-NS sector, this operation projects out the Λ_2 representation (the 2-form field $B_{\mu\nu}$) leaving $\mathbf{S}_y \oplus \Lambda_0$ (i.e. the metric $g_{\mu\nu}$ and the dilaton Φ) as in the unoriented bosonic string. The NS-NS sector then contributes $8 \cdot 9/2 = 36$ physical on-shell massless bosons.

The fermionic sectors, NS-R and R-NS, have the same spectra, and the Ω -projection picks their symmetric combination. The on-shell fermionic massless states form the representations $\mathbf{8}_c \oplus \mathbf{56}_s$ of $Spin(8)$; only *one* massless gravitino survives the Ω -projection. Consistency now requires the spacetime theory to be supersymmetric with *one* MW supercharge in the $\mathbf{16}_s$ of $SO(9, 1)$. Supersymmetry also implies equality in the number of propagating bosonic and fermionic degrees of freedom at all mass levels. To get the counting right, the R-R sector should contribute

²⁰ For the full non-linear form of the self-duality constraint, see Sect. 8.5.

$$(8 + 56) - 36 = 28, \quad (5.53)$$

massless bosonic d.o.f. Thus, out of the three $Spin(8)$ representations in the massless R-R sector of IIB, namely Λ_0 , Λ_2 , and Λ_4^+ , only the Λ_2 should survive the $\Omega = +1$ projection. We may get this result more directly as follows. After the Ω projection, the covariant R-R vertex has the form

$$u^{\alpha\beta} \left(Q_\alpha \tilde{Q}_\beta + \tilde{Q}_\alpha Q_\beta \right) e^{ik \cdot X} \quad (5.54)$$

where

$$Q_\alpha = S_\alpha e^{-\phi/2}, \quad \tilde{Q}_\alpha = \tilde{S}_\alpha e^{-\tilde{\phi}/2} \quad (5.55)$$

are the world-sheet conserved currents associated with the two spacetime supersymmetries of IIB and $u^{\alpha\beta}$ a polarization bispinor. Using the spacetime SUSY algebra

$$\{Q_\alpha, \tilde{Q}_\beta\} = 0, \quad (5.56)$$

we rewrite the vertex in the form

$$u^{\alpha\beta} \left(Q_\alpha \tilde{Q}_\beta - Q_\beta \tilde{Q}_\alpha \right) e^{ik \cdot X}, \quad (5.57)$$

so the polarization satisfies $u^{\alpha\beta} = -u^{\beta\alpha}$. In the representation ring of $Spin(9, 1)$,

$$\mathbf{16}_s \wedge \mathbf{16}_s = \mathbf{120} \equiv \Lambda_3 \equiv \wedge^3 \mathbf{10}. \quad (5.58)$$

Imposing BRST invariance as in Sect. 3.7.3, we get

$$u^{\alpha\beta} \propto k_\mu \varepsilon_{\nu\rho} (\gamma^{\mu\nu\rho})^{\alpha\beta}, \quad \text{with} \quad \left[\begin{array}{l} k^2 = 0, \quad k^\mu \varepsilon_{\mu\nu} = 0, \\ \varepsilon_{\mu\nu} \sim \varepsilon_{\mu\nu} + k_{[\mu} \lambda_{\nu]}, \end{array} \right. \quad (5.59)$$

and the on-shell massless R-R states in the unoriented superstring form the **28** of $SO(8)$ as expected.

This $\Omega = +1$ projection yields the *Type I closed unoriented theory*. “Type I” because a single MW gravitino survives the projection, so the supercharges form *one* copy of the **16**_s of $Spin(9, 1)$. The spacetime spectrum is chiral, and the theory potentially suffers from anomalies. Type IIB was also chiral, but in the closed oriented case, the simple algebraic criterion in the gray box of Sect. 5.1 suffices to guarantee quantum consistency, hence the absence of chiral anomalies.²¹ In the unoriented case, this is not enough, since we have no a priori argument to rule out dangerous tadpoles on $\mathbb{RP}^2$, and we need to check their cancelation explicitly. In Sect. 5.6, we shall see that *the closed unoriented superstring theory by itself is inconsistent* because there is a R-R tadpole which cannot be shifted away. This tadpole arises

²¹ *Ad abundantiam* we shall check absence of anomalies in IIB in Sect. 9.2 by direct computation.

for the same reasons as the dilaton tadpole in the bosonic string; cf. Eq. (4.203), but it is much more dangerous. In the bosonic case, we canceled it by adding open unoriented strings with CP gauge group $SO(2^{13})$. Likewise in the Type I case, we shall need to add open unoriented superstrings with Chan–Paton group $G = SO(2^5)$.

Open Superstring Theories

Finally, we consider the open superstring. We know from the *cartoons* in Sect. 1.1.3 that open strings cannot stand alone; to get a consistent model, we also need a closed string sector. If the closed sector is a Type I or II string, the theory contains massless gravitini, and hence all its sectors—including the open string one—should be supersymmetric. This requires a GSO projection in the open string sector. We remain with two possible open sectors:

$$\text{I:} \quad \text{NS+}, \text{R+} \quad \text{massless sector } \mathbf{8}_v \oplus \mathbf{8}_s, \quad (5.60)$$

$$\text{I':} \quad \text{NS+}, \text{R-} \quad \text{massless sector } \mathbf{8}_v \oplus \mathbf{8}_c. \quad (5.61)$$

As in the bosonic string, we may (and in fact we should) add Chan–Paton degrees of freedom living on the boundaries of the superstring world-sheet. As in the bosonic case, the resulting gauge group will be $U(N)$ in the oriented case and $SO(N)$ or $USp(N)$ in the unoriented one.

BOX 5.4 - Spacetime supercharges in the open string sector

To construct the spacetime supercharge Q_α in the open string sector, we need to understand the spin fields in the presence of a boundary. Mapping the strip to the upper half-plane, and taking into account that world-sheet fermions are half-differentials, we have the boundary conditions

$$\left(\psi^\mu(z) - \eta(\bar{z}) \tilde{\psi}^\mu(\bar{z}) \right) \Big|_{\text{Im } z=0} = 0, \quad \eta(\bar{z}) = \begin{cases} +1 & (\text{NS}) \\ \bar{z}/|z| & (\text{R}). \end{cases}$$

Analogously, the spin fields should also satisfy a linear boundary condition relating the left- and right-moving ones on the boundary, i.e.

$$\left(S_\alpha(z) - P_\alpha{}^\beta \tilde{S}_\beta(\bar{z}) \right) \Big|_{\text{Im } z=0} = 0,$$

for some matrix P to be found. P is determined by consistency with the OPEs

$$\psi^\mu(z) S_\alpha(0) \sim \frac{(\gamma^\mu)_\alpha{}^\beta S_\beta(0)}{z^{1/2}}, \quad \tilde{\psi}^\mu(\bar{z}) \tilde{S}_\alpha(0) \sim \frac{(\gamma^\mu)_\alpha{}^\beta \tilde{S}_\beta(0)}{\bar{z}^{1/2}}$$

which gives

$$\eta(\bar{z}) P \gamma^\mu = \frac{\bar{z}}{|z|} \gamma^\mu P$$

and $P = 1$. We then extend $S_\alpha(z)$ to the full $\mathbb{C}$ using the reflection principle. The bottom line is that we have the same realization of supersymmetry as in the left side of the closed superstring; in particular, has a single supercharge in the **16** of $SO(9, 1)$.

The open sector $\mathbf{8}_s$ or $\mathbf{8}_c$ massless spinors are known as *gaugini* since they are the supersymmetry partners of the (massless) gauge vectors in the $\mathbf{8}_v$ whose vertices in the (-1) picture read

$$\lambda^a e_\mu c(x) \psi^\mu(x) \star e^{-\phi(x)+ik \cdot X(x)} \star, \quad k^2 = k^\mu e_\mu = 0, \quad x \in \partial\Sigma \quad (5.62)$$

with λ^a matrices representing the gauge Lie algebra in its defining representation. Gauge vectors and gaugini transform in the adjoint of the gauge group since SUSY commutes with gauge transformations.

Consistency Conditions

As already anticipated, not all these models are consistent. In particular, as described in **BOX 5.4**, the open sector has only $\mathcal{N} = 1$ SUSY while oriented closed strings have $\mathcal{N} = 2$ local supersymmetry (two MW gravitini) and can be consistently coupled only to “matter” having global $\mathcal{N} = 2$ SUSY. Hence, in the presence of open strings, the closed string sector should be of Type I, that is, *unoriented*. Then the world-sheets are non-oriented, and also the open strings should be *unoriented*. These statements are obvious in terms of spacetime physics; from the viewpoint of the world-sheet theory, the boundary conditions relate each operator $\mathcal{O}$ with its Ω -image $\Omega\mathcal{O}\Omega^{-1}$, effectively inducing a Ω -projection.

The result (taking, say, the first GSO projection (5.60)) is *Type I open plus closed superstring theory* whose massless content, written in terms of transverse $Spin(8)$ representations, is

$$\overbrace{\text{Sy} \oplus \Lambda_0 \oplus \Lambda_2 \oplus (\mathbf{8}_v \otimes \mathfrak{g})}^{\text{bosons}} \oplus \overbrace{\mathbf{56}_s \oplus \mathbf{8}_c + (\mathbf{8}_s \otimes \mathfrak{g})}^{\text{fermions}}, \quad (5.63)$$

where $\mathfrak{g}$ stands for the adjoint representation of the gauge Lie algebra $\mathfrak{so}(N)$ or $\mathfrak{usp}(N)$ (cf. Sect. 3.9). The bosonic massless sector contains the metric $g_{\mu\nu}$, the dilaton Φ , a R-R gauge 2-form $A^{(2)}$, and vectors making one copy of the adjoint of the gauge group. The massless fermions are one MW gravitino, one MW dilatino of opposite chirality, and MW gaugini in the adjoint of $\mathfrak{g}$ with the same chirality as the gravitino.

The fermionic spectrum is chiral, and the theory is potentially affected by gauge/gravitational anomalies. To get a consistent theory, the anomalies should cancel between the gravitational supermultiplet (closed string sector) and the “matter multiplet” $(\mathbf{8}_v \oplus \mathbf{8}_s) \otimes \mathfrak{g}$ from the open strings. Since the first contribution is independent of the gauge group G and the second one depends on G , cancelation may happen only for very special gauge groups G . We will show in Chap. 9 that, in fact, the only anomaly-free gauge group is $SO(32)$. We already mentioned that this is the group for which the dangerous R-R tadpole cancels, see Sect. 5.6 for details.

Conclusion: the methods of the present chapter lead to the construction of three consistent tachyon-free supersymmetric string theories, namely Type IIA, Type IIB,

and Type I with $G = SO(32)$ and two less interesting models 0A and 0B plagued by tachyons. In Chap. 7, we shall construct *additional* nice tachyon-free string theories using a more general and systematic approach.

5.4 2d Fermionic Path Integrals

We wish to check modular invariance for the closed oriented superstrings constructed in Sect. 5.2 by direct computation of their partition functions on the torus. Before going to that, we pause a while to learn how to compute fermionic path integrals on world-sheets of various topologies.

Weyl Fermion on the Torus

We compute the partition function on a torus of periods $(2\pi, 2\pi\tau)$ of a free Weyl fermion $\lambda(z)$ with action

$$\frac{1}{2\pi} \int \bar{\lambda} \bar{\partial} \lambda \quad (5.64)$$

subjected to the general periodicity condition

$$\lambda(w + 2\pi) = e^{\pi i(1-\alpha)} \lambda(w), \quad \lambda(w + 2\pi\tau) = e^{\pi i(1-\beta)} \lambda(w) \quad (5.65)$$

where α, β are real numbers with $-1 < \alpha, \beta \leq 1$.²² The left-moving Fermi-field pair $\lambda(z), \bar{\lambda}(z)$ form a fermionic b, c system with $\lambda = (1 - \lambda) = \frac{1}{2}$ and “non-standard” Hermitian structure, cf. Sect. 2.9.

We see the A -cycle of the torus as the spatial circle, and the B -cycle as periodic Euclidean time. The path integral is given by a trace in the Hilbert space $\mathbf{H}_\alpha$ of Weyl fermions quantized in the circle of unit radius with b.c. the first Eq. (5.65). This twisted b.c. shifts the modes of $\lambda, \bar{\lambda}$ as

$$\lambda(w) = \sum_{m \in \mathbb{Z}} \lambda_{m+(1-\alpha)/2} e^{i[m+(1-\alpha)/2]w} \quad (5.66)$$

so that the raising operators now are

$$\lambda_{-m+(1-\alpha)/2} \quad \text{and} \quad \bar{\lambda}_{-m+(1+\alpha)/2}, \quad m = 1, 2, \dots, \quad (5.67)$$

that is, the appropriate “twisted vacuum” $|\text{tws}\rangle$ satisfies

$$\begin{cases} \lambda_n |\text{tws}\rangle = 0 & \text{for } n \geq -\frac{\alpha}{2} + \frac{1}{2} \\ \bar{\lambda}_n |\text{tws}\rangle = 0 & \text{for } n > \frac{\alpha}{2} - \frac{1}{2} \end{cases} \quad (5.68)$$

²² More generally, with $\alpha \in \mathbb{R}$. Making $\alpha \rightarrow \alpha + 2$ does not modify the periodic boundary conditions of λ , Eq. (5.65), but changes the reference Fermi sea as $|\alpha/2\rangle \rightarrow |\alpha/2 + 1\rangle$ which modifies both the energy level $h - c/24$ and the $U(1)$ charge J . The map $\alpha \rightarrow \alpha + 2$ —which is a non-trivial isomorphism of the operator algebra—is called *spectral flow*.

and hence $|\text{tws}\rangle$ is the Fermi sea $|\alpha/2\rangle$ for $\lambda = 1/2$ as defined in Eq. (2.246). The chiral sea state $|\alpha/2\rangle$ has energy $H = H_L \equiv L_0 - c/24$

$$h - \frac{c}{24} \equiv \frac{1}{2}q(q + Q) - \frac{c}{24}, \quad (5.69)$$

with $Q = 0$, $q = \alpha/2$, and $c = 1$, see Sect. 2.5 especially Eq. (2.256). That is, the H_L -eigenvalue of $|\alpha/2\rangle$ is

$$E_{\alpha/2} \equiv \frac{3\alpha^2 - 1}{24}. \quad (5.70)$$

Exercise 5.3 Check that the formula (5.70) for the “vacuum” energy with b.c. (5.65) agrees with the results of the ζ -function methods in BOX 1.2.

The complex fermion has a $U(1)$ current $:\bar{\lambda}\lambda(z):$ and an associated conserved²³ charge J , described in Sect. 2.5. The reference Fermi sea $|\alpha/2\rangle$ has charge $J = \alpha/2$, while $\lambda(z)$ has charge $+1$ and $\bar{\lambda}(z)$ charge -1 . We define a partition function in the Hilbert space $\mathbf{H}_\alpha$ valued in the $U(1)$ characters as

$$\begin{aligned} Z_\beta^\alpha(\tau) &\stackrel{\text{def}}{=} \text{Tr}_\alpha[q^{L_0 - c/24} e^{\pi i \beta J}] = \\ &= q^{(3\alpha^2 - 1)/24} e^{\pi i \alpha \beta/2} \prod_{m=1}^{\infty} [1 + e^{\pi i \beta} q^{m - (1 - \alpha)/2}] [1 + e^{-\pi i \beta} q^{m - (1 + \alpha)/2}] \\ &= \frac{1}{\eta(q)} \vartheta \left[\begin{matrix} \alpha/2 \\ \beta/2 \end{matrix} \right] (\tau), \end{aligned} \quad (5.71)$$

where in the last line we wrote the answer in terms of the (elliptic) ϑ -function with characteristics using Jacobi’s triple product identity,²⁴ see BOX 5.5 for definitions, notations, and main properties of ϑ -functions.

The world-sheet fermion number is $F \equiv J \bmod 2$, so that $(-1)^F \equiv e^{\pi i J}$. Thus if we have a pair ψ^a ($a = 1, 2$) of MW fermions, which we combine in a complex Weyl fermion $\lambda = (\psi^1 + i\psi^2)/\sqrt{2}$, their torus partition functions with the four possible spin-structures are²⁵

$$(A, A) \quad Z_0^0(\tau) = \text{Tr}_{\text{NS}}[q^{L_0 - c/24}] \quad (5.72)$$

$$(A, P) \quad Z_1^0(\tau) = \text{Tr}_{\text{NS}}[(-1)^F q^{L_0 - c/24}] \quad (5.73)$$

$$(P, A) \quad Z_0^1(\tau) = \text{Tr}_{\text{R}}[q^{L_0 - c/24}] \quad (5.74)$$

$$(P, P) \quad Z_1^1(\tau) = \text{Tr}_{\text{R}}[(-1)^F q^{L_0 - c/24}]. \quad (5.75)$$

²³ For $\lambda = 1/2$ the gravitational anomaly of the $U(1)$ current cancels; cf. Sect. 2.5.

²⁴ We shall show in Sect. 6.3 that the Jacobi’s triple product identity is the mathematical statement of bosonization of massless fermions in 2d.

²⁵ The symbols in the parenthesis specify the fermion b.c. on the torus for each spin-structure; first entry periodicity along the A -cycle, second periodicity along the B -cycle, with A = anti-periodic and P = periodic.

Modular Properties The transformation under the modular group of the Fermi partition functions with given spin-structure, Eqs. (5.72)–(5.75), are read from **BOX 5.5** and the transformation of the η -function (4.57); the results are tabled in **BOX 5.6**.

Dirac Fermion on the Cylinder

We write the cylinder as $[0, \pi] \times \mathbb{R}/2\pi t \mathbb{Z}$ where $t > 0$ is the real modulus of its complex structure. We see the interval $[0, \pi]$ as space and the circle as periodic Euclidean time, i.e. we interpret the path integral as a thermal partition function for the Dirac fermion quantized in the interval. With impose our standard strip b.c.

$$\tilde{\lambda}(\pi, y) = \lambda(\pi, y) \quad \lambda(0, y) = e^{\pi i(1-\alpha)} \tilde{\lambda}(0, y), \quad (5.76)$$

By the doubling trick (Sect. 2.6) the resulting Hilbert space is isomorphic to the one for a Weyl fermion on the full circle $\mathbf{H}_\alpha$. Hence we have

$$\mathrm{Tr}_{\mathbf{H}_\alpha} \left[e^{\beta J} q^{-2\pi t(L_0 - c/24)} \right] = Z_\beta^\alpha(it). \quad (5.77)$$

Crossed-Channel Viewpoint In the crossed channel the path integral is re-interpreted as a closed string of length 2π which propagates for an Euclidean time π/t between suitable boundary states of the fermionic CFT

$$\langle B' | \exp[-\pi(L_0 + \tilde{L}_0 - c/12)/t] | B \rangle. \quad (5.78)$$

Exercise 5.4 Construct the Dirac CFT boundary states $|B\rangle, |B'\rangle$ for given $\alpha, \beta = 0, 1$.

Dirac Fermion on the Klein Bottle

The Klein bottle²⁶ Kl is the quotient of the complex plane $\mathbb{C}$ by a discrete group $G = \langle \xi, \eta \rangle$ of symmetries (acting freely and properly discontinuously) which does not preserve orientation:

$$\xi: z \mapsto z + 2\pi, \quad \eta: z \mapsto -\bar{z} + 2\pi it. \quad (5.79)$$

We have different models of the Klein bottle corresponding to different choices of the fundamental domain²⁷ for G acting in $\mathbb{C}$. See Fig. 5.1. As a fundamental domain in the complex plane of coordinate $z \equiv x + iy$, we can either take the rectangle

$$R \equiv \{0 \leq x \leq 2\pi, 0 \leq y \leq 2\pi t\}, \quad (5.80)$$

with boundary identifications

²⁶ See also **BOX 1.5**.

²⁷ Cf. **Definition 4.1**.

BOX 5.5 - ϑ -functions with characteristics

The ϑ -function with characteristics (a, b) on a torus of modulus $\tau \in \mathcal{H}$ is the convergent sum

$$\vartheta \begin{bmatrix} a \\ b \end{bmatrix} (z|\tau) \stackrel{\text{def}}{=} \sum_{k \in \mathbb{Z}} \exp[\pi i(k+a)^2 \tau + 2\pi i(k+a)(z+b)]$$

It satisfies the identities (here $m, n \in \mathbb{Z}$)

$$\begin{aligned} \vartheta \begin{bmatrix} a \\ b \end{bmatrix} (z+m+n\tau|\tau) &= e^{2\pi i m a - i\pi n^2 \tau - 2\pi i n(z+b)} \vartheta \begin{bmatrix} a \\ b \end{bmatrix} (z|\tau) \\ \vartheta \begin{bmatrix} a \\ b \end{bmatrix} (z|\tau) &= e^{i\pi a^2 \tau + 2\pi i a(z+b)} \vartheta \begin{bmatrix} 0 \\ 0 \end{bmatrix} (z+a\tau+b|\tau) \\ \vartheta \begin{bmatrix} a+m \\ b+n \end{bmatrix} (z|\tau) &= e^{2\pi i n a} \vartheta \begin{bmatrix} a \\ b \end{bmatrix} (z|\tau) \quad \vartheta \begin{bmatrix} -a \\ -b \end{bmatrix} (z|\tau) = \vartheta \begin{bmatrix} a \\ b \end{bmatrix} (-z|\tau) \end{aligned}$$

The dependence on the first argument, z , may be absorbed in the characteristics

$$\vartheta \begin{bmatrix} a \\ b \end{bmatrix} (z|\tau) = \vartheta \begin{bmatrix} a \\ b+z \end{bmatrix} (0|\tau)$$

and we write $\vartheta \begin{bmatrix} a \\ b \end{bmatrix} (\tau)$ for $\vartheta \begin{bmatrix} a \\ b \end{bmatrix} (0|\tau)$ (the so-called *theta-constants*).

The *Jacobi triple product identity* allows to rewrite the function as an infinite product

$$\vartheta \begin{bmatrix} a \\ b \end{bmatrix} (\tau) = \eta(\tau) e^{2\pi i a b} q^{a^2/2 - 1/24} \prod_{n=1}^{\infty} (1 + q^{n+a-1/2} e^{2\pi i b}) (1 + q^{n-a-1/2} e^{-2\pi i b}),$$

Its modular transformations are (for $|\arg \sqrt{-i\tau}| < \pi/2$)

$$\vartheta \begin{bmatrix} a \\ b \end{bmatrix} (\tau+1) = e^{-\pi i(a^2-a)} \vartheta \begin{bmatrix} a \\ a+b-\frac{1}{2} \end{bmatrix} (\tau), \quad \vartheta \begin{bmatrix} a \\ b \end{bmatrix} (-1/\tau) = \sqrt{-i\tau} e^{2\pi i a b} \vartheta \begin{bmatrix} -b \\ a \end{bmatrix} (\tau),$$

ϑ 's with characteristics $a, b = 0, \frac{1}{2}$ are related to *spin-structures* on T^2 . They have special names

$$\vartheta \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} = \vartheta_1, \quad \vartheta \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} = \vartheta_2, \quad \vartheta \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \vartheta_3, \quad \vartheta \begin{bmatrix} 0 \\ 1/2 \end{bmatrix} = \vartheta_4$$

Another widely used notation for them is

$$\vartheta_{ab}(\tau) = \vartheta \begin{bmatrix} a/2 \\ b/2 \end{bmatrix}, \quad a, b = 0, 1$$

From the definition $\vartheta_{11}(\tau) = 0$, we have Jacobi's "abstruse identity" (see Sect. 6.3 for a proof)

$$\vartheta_{00}(\tau)^4 - \vartheta_{01}(\tau)^4 - \vartheta_{10}(\tau)^4 = 0$$

A useful identity is

$$\partial_z \vartheta_{11}(z|\tau) \Big|_{z=0} = -2\pi \eta(\tau)^3 \quad \clubsuit$$

where $\eta(\tau)$ is the Dedekind function $\eta(\tau) = q^{1/24} \prod_{m=1}^{\infty} (1 - q^m)$ ($q \equiv e^{2\pi i \tau}$) and

$$\vartheta_{11}(z|\tau) \equiv -2 e^{\pi i \tau/4} \sin(\pi z) \prod_{m=1}^{\infty} (1 - q^m) (1 - 2 \cos(\pi z) q^m + q^{2m})$$

BOX 5.6 - Modular properties of Z^α_β

Using modular transformation formulae from BOX 5.5 and Eq. (4.57)

$$\frac{1}{\eta(\tau+1)} \vartheta \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] (\tau+1) = \frac{e^{-i\pi/12}}{\eta(\tau)} e^{-\pi i(a^2-a)} \vartheta \left[\begin{smallmatrix} a \\ a+b-\frac{1}{2} \end{smallmatrix} \right] (\tau)$$

$$\frac{1}{\eta(-1/\tau)} \vartheta \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] (-1/\tau) = \frac{1}{\sqrt{-i\tau} \eta(\tau)} \sqrt{-i\tau} e^{2\pi iab} \vartheta \left[\begin{smallmatrix} -b \\ a \end{smallmatrix} \right] (\tau) \equiv \frac{e^{2\pi iab}}{\eta(\tau)} \vartheta \left[\begin{smallmatrix} -b \\ a \end{smallmatrix} \right] (\tau) \quad (\diamond)$$

In particular

$$\begin{aligned} Z^0_0(\tau+1) &= e^{-i\pi/12} Z^0_1(\tau), & Z^1_0(\tau+1) &= e^{i\pi/6} Z^1_0(\tau), & Z^0_1(\tau+1) &= e^{-i\pi/12} Z^0_0(\tau) \\ Z^0_0(-1/\tau) &= Z^0_0(\tau), & Z^1_0(-1/\tau) &= Z^0_1(\tau), & Z^0_1(-1/\tau) &= Z^0_1(\tau) \equiv Z^1_0(\tau) \\ Z^1_1(\tau+1) &= e^{i\pi/6} Z^1_1(\tau), & & & Z^1_1(-1/\tau) &= e^{i\pi/2} Z^1_1(\tau) \end{aligned}$$

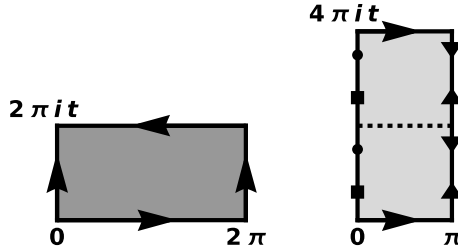


Fig. 5.1 Two fundamental domains for the K1 group G acting on $\mathbb{C}$ (we set $t = 1/2$). The region R is dark gray, while the region $R^\vee$ light gray. Boundary segments are identified as in the main text (arrows specify the orientation of identified boundaries; same symbol points are identified)

$$(0, y) \sim (2\pi, y) \quad 0 \leq y \leq 2\pi t \quad (5.81)$$

$$(x, 0) \sim (2\pi - x, 2\pi t) \quad 0 \leq x \leq 2\pi. \quad (5.82)$$

or the dual rectangle

$$R^\vee \equiv \{0 \leq x \leq \pi, 0 \leq y \leq 4\pi t\}, \quad (5.83)$$

with boundary identifications

$$(0, y) \sim (0, y + 2\pi t) \quad (\text{crosscap}) \quad (5.84)$$

$$(\pi, y) \sim (\pi, y + 2\pi t) \quad (\text{crosscap}) \quad (5.85)$$

$$(x, 0) \sim (x, 4\pi t) \quad 0 \leq x \leq \pi. \quad (5.86)$$

We call R and $R^\vee$ the direct-channel and crossed-channel models of K1, respectively.

The direct-channel model presents the Klein bottle of real modulus $t > 0$ as obtained from a cylinder of circumference 2π and length $2\pi t$ by identifying the two boundaries with a Ω -twist. The Klein bottle partition function then differs from the torus one with $\tau = it$ just by the insertion of Ω in the Hilbert space trace.

It remains to compute the traces

$$\text{Tr}_{\alpha, \tilde{\alpha}} \left[\Omega e^{\pi i (\beta F + \tilde{\beta} \tilde{F})} q^{L_0 - c/24} \bar{q}^{\tilde{L}_0 - c/24} \right], \quad q = e^{-2\pi t}, \quad (5.87)$$

for $\alpha, \tilde{\alpha}, \beta, \tilde{\beta} = 0, 1$. Since Ω interchanges left- and right movers, the amplitude vanishes if $\alpha \neq \tilde{\alpha}$ while the states that contribute have $F \equiv \tilde{F}$ and $L_0 \equiv \tilde{L}_0$, and the expression reduces to

$$\text{Tr}_{\alpha, \alpha} \left[\Omega e^{\pi i (\beta + \tilde{\beta}) F} e^{-4\pi t (L_0 - c/24)} \right] = Z_{\beta + \tilde{\beta}}^{\alpha}(2it). \quad (5.88)$$

Crossed-Channel Viewpoint The crossed-channel model represents K1 as a cylinder in the dual channel of circumference $4\pi t$ and length π which is closed at both ends by crosscaps. The K1 partition function takes the form

$$\langle C' | e^{-\pi/t (L_0 + \tilde{L}_0 - 2)} | C \rangle \quad (5.89)$$

for suitable Fermi crosscap states $|C\rangle, |C'\rangle$.

Exercise 5.5 Construct the Dirac CFT crosscap states $|C\rangle, |C'\rangle$ for given $\alpha, \beta = 0, 1$.

Dirac Fermion in the Möbius Strip

Again we have two distinct models of the Möbius strip which correspond to the direct- and crossed channel, respectively. The direct model has fundamental domain and boundary identifications ($z = x + iy$)

$$R = \{0 \leq x \leq \pi, 0 \leq y \leq 2\pi t: (x, 0) \sim (\pi - x, 2\pi t)\} \quad (5.90)$$

where $t > 0$ is the real modulus. The direct model sees the Möbius surface as a strip of width π and length $2\pi t$ where the two ends are glued together with an orientation flip Ω . Hence the partition function is interpreted as a trace over the NS or R sector of the Hilbert space of the complex fermion quantized in the strip $[0, \pi]$ with the usual b.c. (5.76) and the insertion of Ω

$$\text{Tr}_{\alpha} \left[\Omega e^{\pi i \beta J} e^{-2\pi t (L_0 - c/24)} \right]. \quad (5.91)$$

In the open sector Ω acts as

$$\Omega \lambda(w) \Omega^{-1} = \tilde{\lambda}(\pi - \bar{w}) = \lambda(w - \pi), \quad (5.92)$$

BOX 5.7 - Ω -twisted fermionic partition functions

For a complex fermionic fields λ , we have

$$\mathrm{Tr}_R[\Omega e^{-2\pi t(L_0 - c/24)}] = 2q^{1/8-1/24} \left(\prod_{m=1}^{\infty} \sum_{N_n=0}^1 (-e^{-2\pi t})^{nN_n} \right)^2 = 2q^{1/12} \prod_{n=1}^{\infty} (1 + q^{2m})^2 (1 - q^{2m-1})^2$$

From Eq. (5.71), we have

$$4 \quad \prod_{m=1}^{\infty} (1 - q^{m-1/2})^2 = q^{1/24} Z_1^0(q), \quad \prod_{m=1}^{\infty} (1 + q^m)^2 = \frac{1}{2} q^{-1/12} Z_1^1(q).$$

so

$$\mathrm{Tr}_R[\Omega e^{-2\pi t(L_0 - c/24)}] = Z_1^0(q^2) Z_1^1(q^2).$$

Likewise

$$\mathrm{Tr}_R[\Omega (-1)^F e^{-2\pi t(L_0 - c/24)}] = 2q^{1/12} \prod_{m \geq 1} (1 - q^{2m})^2 (1 + q^{2m-1})^2 = 2\eta(q^2)^2 Z_0^0(q^2)$$

where in the last equality we used the Schwarz reflection principle (“doubling trick”). In terms of modes, this is

$$\Omega \lambda_r \Omega^{-1} = e^{-\pi i r} \lambda_r. \quad (5.93)$$

In the NS sector, the phase is imaginary and squares to -1 . In the sum over states in (5.91), we have to insert the phases (5.93); their net effect is a shift in the characteristics of the ϑ -functions. The partition functions (5.91) for $\alpha = 1$ (R sector) and $\beta = 0, 1$ are computed in BOX 5.7.

Exercise 5.6 Compute the partition functions (5.91) for the NS sector.

Crossed-Channel Viewpoint The crossed-channel model of the Möbius has fundamental domain and boundary identifications

$$R^\vee = \{0 \leq x \leq \pi/2, 0 \leq y \leq 4\pi t : (x, 0) \sim (x, 4\pi t), (\pi/2, y) \sim (\pi/2, y + 2\pi t)\}, \quad (5.94)$$

i.e. we have a cylinder of circumference $4\pi t$ and length $\pi/2$ with a crosscap glued in the boundary at $x = \pi/2$ while at $x = 0$ we have an ordinary Neumann boundary.

Inserting $\Omega e^{\pi i \beta J}$ in the Hilbert space trace is equivalent to performing the path integral over fields with the periodicity condition

$$\lambda(w + 2\pi i t) = -e^{\pi i \beta} \lambda(w - \pi), \quad (5.95)$$

so (specializing to $\beta = 0, 1$)

$$\lambda(w + 4\pi i t) = -e^{\pi i \beta} \lambda(w + 2\pi i t - \pi) = e^{2\pi i \beta} \lambda(w - 2\pi) \equiv \lambda(w - 2\pi) = \mp \lambda(w) \quad (5.96)$$

where the upper (lower) sign is for the open NS (resp. R) sector. Hence in the R sector of the open channel, the fields are periodic in the dual channel of period $4\pi t$, thus corresponding to the exchange of R-R closed states in the dual channel, while the open NS sector corresponds to a NS-NS exchange in the crossed channel.

5.5 Modular Invariance in Type II

In this section, we check modular invariance of the closed oriented Type II superstrings—already established in Sect. 5.2 from first principles—by direct computation of their one-loop amplitude. To simplify the formulae, we set $\alpha' = 2$; when needed, the dependence on α' may be easily restored by dimensional analysis.

Arguing exactly as we did for the bosonic string (see Sect. 4.6), we conclude that the torus amplitude Z_{T^2} is given, in terms of the physical particle spectrum, by the same Coleman–Weinberg (CW) formula [14] which holds in QFT except that the region of integration over the Schwinger parameters τ_1, τ_2 should be restricted from the strip region $R \subset \mathcal{H}$ in the upper half-plane to the fundamental domain $F_0 \simeq \mathcal{H}/PSL(2, \mathbb{Z})$ of the moduli space of tori, see Fig. 4.3. Then we have

$$Z_{T^2} = V_{10} \int_{F_0} \frac{d^2 \tau}{\tau_2} \int \frac{d^{10} k}{(2\pi)^{10}} \sum_i (-1)^{\mathbf{F}_i} q^{(k^2 + m_i^2)/2} \bar{q}^{(k^2 + \tilde{m}_i^2)/2}, \quad (5.97)$$

where V_{10} is the volume of 10d spacetime, $q = \exp(2\pi i \tau)$, and:

- $\sum_i$ stands for the trace over the space of physical states at fixed momentum k_μ , isomorphic to the trace over the Hilbert space $\mathbf{H}_\perp$ of *transverse* oscillators. The trace includes a sum over the different $(\alpha, F; \tilde{\alpha}, \tilde{F})$ sectors of $\mathbf{H}_\perp$;
- spacetime fermions have a *minus sign* in the CW formula. Here $\mathbf{F}_i$ is the spacetime fermion number, not to be confused with the world-sheet one F ;
- the masses are expressed in terms of the transverse level numbers, $N_\perp$ and $\tilde{N}_\perp$

$$m^2 = 2(N_\perp - \nu), \quad \tilde{m}^2 = 2(\tilde{N}_\perp - \tilde{\nu}), \quad (5.98)$$

with $\nu \equiv \frac{1}{2}(1 - \alpha)$, $\tilde{\nu} \equiv \frac{1}{2}(1 - \tilde{\alpha})$ equal to 0, $\frac{1}{2}$ in the R, NS sectors, respectively.

Exercise 5.7 Deduce Eq. (5.97) in the covariant gauge from string first principles.

In each sector $(\alpha, F; \tilde{\alpha}, \tilde{F})$ the trace over the transverse oscillators, including the integral over the transverse bosonic zero-mode $k_\perp$, decouples into the product of independent traces over the Hilbert space of each transverse field X^i , ψ^i and $\tilde{\psi}^i$, that

is, in the product of the corresponding free-field torus partition functions. We have already computed all the relevant path integrals in Sects. 4.1.1 and 5.4.

The Partition Function of X

The path integral for a single non-compact scalar field X was studied in Sect. 4.1.1, see Eq. (4.55). The total contribution from the oscillators of X , together with the integral over its zero-mode (momentum integral), is²⁸

$$Z_X(\tau) = (8\pi^2\tau_2)^{-1/2} |\eta(q)|^{-2} \quad (5.99)$$

where, as always, $\eta(q)$ is the Dedekind function. In Eq. (5.97) there is no contribution from the two longitudinal $X^\pm$ oscillators. However, their zero-modes (k_+ , k_-) do contribute, giving an additional factor²⁹ $(8\pi^2\tau_2)^{-1}$.

The Partition Function of ψ

The partition function on the left-moving fermions depends on their spatial periodicity specified by $\alpha \in \{0, 1\}$ ($\alpha = 0$ NS, $\alpha = 1$ R sector), and includes inside the trace the GSO projection operator

$$P_\pm = \frac{1}{2} [1 \pm (-1)^{F_{\text{GSO}}}] \quad (5.100)$$

on the appropriate chirality selected by the $\text{GSO}\pm$ projection. As in Sect. 2.9 we replace the eight transverse MW fermions ψ^i ($i = 1, \dots, 8$) by four complex Weyl fermions λ^j ($j = 1, \dots, 4$). The partition function of a single free Weyl fermion λ , subjected to the general periodicity condition (5.65), is given by Eq. (5.71).

GSO Projection We compute the chiral partition function $Z_\psi^\pm(\tau)$ of the superstring 8 real transverse fermions ψ^i subjected to the GSO projection which keeps the sectors NS+ and R $\pm$. Comparing with Sect. 3.1 we see that the Fermi number F_{GSO} relevant for the GSO projection differs from the F used in Eqs. (5.72)–(5.75) by the Fermi numbers (mod 2) of the longitudinal $\psi^\pm$ zero-modes and of the spinor ghosts β, γ . Using the standard (-1) picture³⁰ for the NS sector we see that in the NS sector

$$(-1)^{F_{\text{GSO}}} = -(-1)^F, \quad (5.101)$$

while (by definition) in the R sector, $(-1)^{F_{\text{GSO}}} = \pm(-1)^F$ if R $\pm$ survives. Then³¹

²⁸ Times an overall length factor; the product of all these length factors over all directions produces the V_{10} in front of Eq. (5.97).

²⁹ If the target space is Lorentzian (as contrasted to Euclidean) there is an extra overall factor i since the k_0^2 term has the “wrong” sign and must be Wick rotated.

³⁰ “Standard picture” corresponds to the ghosts’ sea of OCQ; cf. Sect. 3.3. In the text we are implicitly using the isomorphism between the OCQ and light-cone Hilbert spaces. More in general, the covariant chirality operator is $(-1)^{F_{\text{GSO}}} \equiv (-1)^{t \cdot \lambda}$, with λ the $SO(10, 2)$ weight; cf. Eq. (3.11).

³¹ We stress that due to the “wrong sign” projection (5.101) in the (which projects put the identity), $Z_\psi^\pm$ is not the partition function of an algebra $\mathfrak{A}$ hence its modular transformations are not given by Eq. (5.36) or rather are given by that expression only up to signs.

$$\begin{aligned}
Z_{\psi}^{\pm} &= \text{Tr}_{\text{NS}} \left[\frac{1+(-1)^{F_{\text{GSO}}}}{2} q^{L_0-c/24} \right] - \text{Tr}_{\text{R}} \left[\frac{1\pm(-1)^{F_{\text{GSO}}}}{2} q^{L_0-c/24} \right] = \\
&= \frac{1}{2} \left((Z^0_0(\tau))^4 - (Z^0_1(\tau))^4 - (Z^1_0(\tau))^4 \mp (Z^1_1(\tau))^4 \right),
\end{aligned} \tag{5.102}$$

where the minus in the second term of the first line arises from the spacetime fermion number sign factor $(-1)^{F_i}$ in the CW formula (5.97). The partition functions for the right-movers $\tilde{\psi}$ are the complex conjugate of the left-movers ones, i.e. $(Z_{\psi}^{\pm})^*$.

Modular Properties

Putting everything together, the closed superstring one-loop vacuum amplitude is

$$Z_{T^2} = V_{10} \int_{F_0} \frac{d^2\tau}{32\pi^2\tau_2^2} Z_X(\tau, \bar{\tau})^8 Z_{\psi}^+(\tau) Z_{\psi}^{\pm}(\tau)^*, \quad \begin{cases} + & \text{for IIB} \\ - & \text{for IIA.} \end{cases} \tag{5.103}$$

As in the bosonic string, modular invariance of the integrand is a necessary consistency condition.³² $d^2\tau/\tau_2^2$ is the $SL(2, \mathbb{R})$ -invariant Poincaré volume form (BOX 4.5) which is obviously modular invariant, as it is $Z_X(\tau)$; cf. Eq. (4.58). It remains to discuss the modular properties of the GSO-projected fermionic traces $Z_{\psi}^{\pm}$ and $(Z_{\psi}^{\pm})^*$.

Modular Properties of Fermi Partitions Functions

The modular transformations of the functions $Z_{\beta}^{\alpha}(\tau)$ are given in BOX 5.6. There is a subtlety in the BOX which requires a comment. The partition function $Z^1_1(\tau)$ *vanishes identically*: in the path integral formalism this is due to the presence of a Fermi zero-mode for periodic b.c., while in the operator language, it is due to the double-degeneracy of the Ramond ground state which for a Weyl fermion λ is a 2d spinor with two components of chirality $+1$ and -1 . The insertion of $(-1)^F$ makes the towers of states constructed by acting with oscillators on these two R vacua to contribute with opposite signs, producing a total cancelation. In yet another language, $Z_X(\tau)^2 |Z^1_1(\tau)|^2$ is the Witten index of a $(2, 2)$ free massless chiral supermultiplet, thus is τ independent and in fact zero. Since $Z^1_1(\tau)$ is zero, it is modular invariant for all choices of overall phases in its transformation; we *declare* these phases to be as in the last line of BOX 5.6. The physical significance of this statement will be explained in Remark 5.1 below.

From BOX 5.6, we see that the transformation $S: \tau \mapsto -1/\tau$ acts on the set

$$\left\{ (Z^0_0)^4, (Z^1_0)^4, (Z^0_1)^4, (Z^1_1)^4 \right\}, \tag{5.104}$$

by permuting the second and third elements, leaving invariant the expression

$$2 Z_{\psi}^{\pm} \equiv (Z^0_0)^4 - (Z^1_0)^4 - (Z^0_1)^4 \mp (Z^1_1)^4, \tag{5.105}$$

so the partition functions $Z_{\psi}^{\pm}(\tau)$ in Eq. (5.102) are S -invariant.

³² In particular it is required to justify restriction of the integration domain to F_0 .

Since $(e^{-i\pi/12})^4 \equiv -e^{2\pi i/3}$, under $T: \tau \rightarrow \tau + 1$, we have

$$\begin{aligned} 2 Z_{\psi}^{\pm}(\tau + 1) &= Z_0^0(\tau + 1)^4 - Z_1^0(\tau + 1)^4 - Z_0^1(\tau + 1)^4 \mp Z_1^1(\tau + 1)^4 \\ &= -e^{2\pi i/3} Z_1^0(\tau)^4 + e^{2\pi i/3} Z_0^0(\tau)^4 - e^{2\pi i/3} Z_1^0(\tau)^4 \mp e^{2\pi i/3} Z_1^1(\tau)^4 \quad (5.106) \\ &= 2 e^{2\pi i/3} Z_{\psi}^{\pm}(\tau). \end{aligned}$$

The two combinations $Z_{\psi}^+(\tau) Z_{\psi}^{\pm}(\tau)^*$ are thus fully modular invariant, and hence the integrand of the torus partition function (5.103) is *modular invariant*, and we should restrict the integral to the fundamental domain F_0 to avoid multiple counting.

Discussion A number of *fundamental* remarks are in order:

Remark 5.1 The one-loop vacuum amplitude in a supersymmetric theory is expected to vanish by cancelations between fermions and bosons. This holds because of $Z_1^1(\tau) = 0$ and one identity found by Jacobi and called by him “*aequatio identica satis abstrusa*” (the “abstruse identity”)

$$\vartheta_3(q)^4 = \vartheta_4(q)^4 + \vartheta_2(q)^4 \Rightarrow (Z_0^0)^4 - (Z_1^0)^4 - (Z_1^1)^4 = 0. \quad (5.107)$$

We shall prove a more general version of this identity (due to Riemann) in Sect. 6.3.

Remark 5.2 At the end of Sect. 5.1, we stated that global $\mathbf{Diff}^+$ anomalies cancel in all perturbative amplitudes, to all loop orders, provided *all* BRST-invariant amplitudes at genus 1 are modular invariant. Above we have shown modular invariance of the one-loop amplitude *without* operator insertions. We need to check that modular invariance is not spoiled by the insertion of *arbitrary* GSO-allowed BRST-invariant vertex operators. This is easy; consider the left-moving torus amplitude

$$\left\langle \int \mathcal{O}_1(z_1) dz_1 \cdots \int \mathcal{O}_s(z_s) dz_s \right\rangle_{\beta}^{\alpha} \quad (5.108)$$

where $\alpha, \beta = 0, 1$ label the four spin-structures, as always. The integrated BRST-invariant operators $\mathcal{O}_i(z_i)$ have weight $h_i = 1$ and commute with $(-1)^{F_{\text{GSO}}}$ by the GSO projection. We claim that the above amplitudes transform under the modular group as the partition functions Z_{β}^{α} for the same spin-structure. This is obvious if the $\mathcal{O}_i(z_i)$ are NS vertices, that is, polynomials in ψ^{μ} , their derivatives, and the other left-moving 2d fields. The path integral which computes (5.108) is then Gaussian, and the amplitude has the schematic form³³

$$\frac{\left(\begin{array}{c} \text{fermionic determinant} \\ \text{with spin-structure } \alpha, \beta \end{array} \right)}{\left(\begin{array}{c} \text{bosonic determinant} \\ \text{with spin-structure } \alpha, \beta \end{array} \right)} \times \int dz_1 \cdots dz_s \left(\begin{array}{c} \text{Wick contractions of} \\ \text{fields in } \mathcal{O}_1, \dots, \mathcal{O}_s \end{array} \right) \quad (5.109)$$

³³ In the presence of zero-modes, i.e. when $\alpha = \beta = 1$ the determinants are replaced by the *primed* determinants with the zero eigenvalues omitted, and the amplitude is non-zero only if there are enough Fermi-field insertions to soak up all zero-modes.

This amplitude transforms under the modular group in the classical way *times* a quantum phase arising from the **Diff**⁺ anomaly. The only potential source of phase ambiguity is the fermionic determinant, which for $(\alpha, \beta) \neq (1, 1)$ is just the fermionic partition function in the (α, β) spin-structure. Hence the anomalous phases are the same ones with and without insertions. Closure of OPE implies that the same statement holds for R vertices; a non-zero amplitude contains an even number of spin fields, and we may replace each pair $S_\alpha(z_1) S_\beta(z_2)$ by its exact OPE expansion

$$S_\alpha(z_1) S_\beta(z_2) = \sum_{\alpha} f_{\alpha}(z_1 - z_2) O_{\alpha}(z_2) \quad (5.110)$$

whose coefficients $O_{\alpha}(z)$ are NS operators. The case $(\alpha, \beta) = (1, 1)$ is special because of the zero-mode; we need to introduce Fermi fields to absorb the zero-modes. For a single Weyl fermion, we have

$$\begin{aligned} \langle : \bar{\lambda} \lambda : \rangle_1^1 &= \int_A \frac{dz}{2\pi} \langle : \bar{\lambda} \lambda(z) : \rangle_1^1 = \frac{1}{2\pi} \text{Tr}_R [(-1)^F F q^{L_0 - c/24}] = \\ &= \frac{1}{4\pi^2 i} \partial_z \vartheta_1(z; \tau) \Big|_{z=0} = \frac{1}{2\pi i} \eta(\tau)^2 \end{aligned} \quad (5.111)$$

(for the last equality see **BOX 5.5**). Under T , this amplitude gets multiplied by the anomalous phase $e^{\pi i/6}$, while under S , a part for the classical factor $-\tau$, it picks up the anomalous phase $e^{\pi i/2}$. These anomalous phases agree with the ones in **BOX 5.6**. We stress that in the presence of general vertex insertions, the $\alpha = \beta = 1$ path integral will no longer vanish, nor will the sum of the other three.

Remark 5.3 The argument in the previous *Remark* exploits the fact that the world-sheet theory is free. However the result holds *for all* $(1, 1)$ SCFT on Σ with the correct central charges since the anomalous phases are universal (final arguments of Sect. 5.1).

Remark 5.4 A general one-loop amplitude has the form

$$\mathcal{A} = \int_{F_0} \frac{d^2 \tau}{\tau_2^2} F(\tau, \bar{\tau}), \quad (5.112)$$

for some real-analytic function $F(\tau, \bar{\tau})$ such that

$$F\left(\frac{a\tau+b}{c\tau+d}, \frac{a\bar{\tau}+b}{c\bar{\tau}+d}\right) = F(\tau, \bar{\tau}) \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}) \quad (5.113)$$

The fundamental domain F_0 has finite Poincaré volume

$$\int_{F_0} \frac{d^2 \tau}{\tau_2^2} = \frac{2\pi}{3}. \quad (5.114)$$

thus if $F(\tau, \bar{\tau})$ is bounded, the amplitude $\mathcal{A}$ is finite. F_0 is biholomorphic to a punctured sphere, where the puncture is at $\tau = i\infty$, i.e. $q \rightarrow 0$. The integrand $F(\tau, \bar{\tau})$ may possibly diverge only in the $q \rightarrow 0$ limit which is controlled by the *lightest* state $F = O((q\bar{q})^{m_{\text{lightest}}^2})$. In absence of tachyons the integrand is bounded and therefore the amplitude is *finite*. The argument can be generalized to all loop orders:

Type II superstring is perturbatively consistent and *finite*

The situation is better than in most QFTs. The perturbative series itself is however only asymptotic, i.e. the theory *does have* interesting non-perturbative phenomena.

Remark 5.5 Modular invariance forces the signs in Eq. (5.102) to be as written; they are exactly the *twist by extra signs* predicted on general grounds in Note 5.1: Z_β^α should carry the extra sign $(-1)^{\alpha+\beta}$. This twist fixes the relative sign between $(Z_0^0)^4$ and $(Z_1^1)^4$; this means that modular invariance requires the states of the R sector to have Fermi statistics. Thus *modular invariance implies the spacetime Spin & Statistics Theorem*. We proved it by a different (but related) argument in BOX 5.3.

Remark 5.6 From our analysis, we see that to have both modular invariance and mixed R-NS/NS-R sectors the number of (transverse) fermions must be a multiple of *eight*. Indeed invariance under $\tau \mapsto \tau + 1$ requires $L_0 - \tilde{L}_0 \in \mathbb{Z}$ for all states. For one real fermion, the difference in ground state energies in the R-NS sector is (BOX 1.2)

$$\frac{1}{24} - \left(-\frac{1}{48}\right) = \frac{1}{16} \quad (5.115)$$

For eight fermions, this is $\frac{1}{2}$, so states with an odd number of transverse NS excitations (as required by GSO) are level-matched. This also follows from the requirement that the (transverse) spin fields are local with respect to themselves; cf. Sect. 2.9.

Remark 5.7 The argument in the previous *Remark* can be replaced by the ones at the end of Sect. 5.1. The partition function of the left-moving fermions should pick up a 6-th root of unity under T , so $c = 4n$, i.e. $8n$ MW fermions.

Remark 5.8 In *Type 0*, superstrings modular invariance is implemented by a different mechanism; the fermionic trace is

$$\frac{1}{2} \left[|Z_0^0(\tau)|^n + |Z_1^0(\tau)|^n + |Z_0^1(\tau)|^n \mp |Z_1^1(\tau)|^n \right] \quad (5.116)$$

with $n = 8$. This the *diagonal modular invariant*, and is modular invariant for all $n \in \mathbb{N}$, since the phases cancel in the absolute values, while S and T permute the three non-zero terms in (5.116).

Fig. 5.2 The world-sheet of the cylinder amplitude. On the two boundaries we have Chan-Paton labels i and j



5.6 Divergences and Tadpoles in Type I Theories

We claimed that Type I is consistent only when the Chan–Paton gauge group is $SO(32)$. There are various ways of seeing this. The main tool to detect inconsistencies is to require the absence of divergences in the open string one-loop amplitudes, or, equivalently, require the absence of tree-level tadpoles which cannot be consistently shifted away, i.e. canceled by a redefinition of the vacuum on which we define the theory. This criterion plays the same role for open strings as modular invariance in the closed case.

Cylinder Amplitude

Let us pretend for a moment that there is an *oriented* open superstring theory. We already know that such a model is inconsistent, but the argument in Sect. 5.3 was based on spacetime low-energy physics rather than stringy first principles. Now we wish to see how the inconsistency is reflected at the full superstring level in terms of one-loop divergences/disk tadpoles which spoil the validity of the perturbative theory.

Note 5.2 In Sect. 3.8.1, we showed that the absence of tadpoles follows from the Ward identities of spacetime supersymmetry. Hence non-zero tadpoles imply a violation of supersymmetry, as found (from a different viewpoint) in Sect. 5.3.

The open string one-loop processes are given by cylinder amplitudes, see Fig. 5.2. The open string sectors are labeled by $\alpha = 0, 1$ (NS vs. R) and by the Chan–Paton labels (i, j) , $i, j = 1, 2, \dots, N$, subjected to the GSO projection,

$$P_{\text{GSO}} = \frac{1}{2} \sum_{\beta=0}^1 (-1)^{\beta F_{\text{GSO}}}. \quad (5.117)$$

Then the open one-loop vacuum amplitude takes the form³⁴

³⁴ Tr'_{α} is the trace over the α sector of the open string with the zero-modes of X^{μ} omitted.

$$\begin{aligned}
Z_{\text{Cy}} &= \frac{N^2}{2} \sum_{\alpha, \beta=0}^1 (-1)^\alpha \int_0^\infty \frac{dt}{4t} \int \frac{d^{10}k}{(2\pi)^{10}} e^{-4\pi t k^2} \text{Tr}'_\alpha \left[(-1)^{\beta F} e^{-2\pi t (L_0 - c/24)} \right] \\
&= N^2 \int_0^\infty \frac{dt}{8t} (16\pi^2 t)^{-5} \eta(it)^{-8} [Z_0^0(it)^4 - Z_0^1(it)^4 - Z_1^0(it)^4 - Z_1^1(it)^4]
\end{aligned} \tag{5.118}$$

where t is the real modulus of the cylinder, i.e. the radius of the circle (the length of the cylinder is fixed to π). Of course, Z_{Cy} vanishes by spacetime supersymmetry; the spacetime fermionic contribution, $\alpha = 1$, exactly cancel the bosonic one, $\alpha = 0$. Indeed, the RHS of (5.118) vanishes by $Z_1^1 = 0$ and the abstruse identity (5.107).

Z_{Cy} may be interpreted in the crossed channel as a *tree-level* closed string process; a closed string propagates between the two boundary states associated with the open string boundary condition.³⁵ In the closed string channel, $\beta = 0$ corresponds to the NS-NS sector and $\beta = 1$ to the R-R one. We write

$$Z_{\text{Cy}} = Z_0 - Z_1 \tag{5.119}$$

$$Z_0 = N^2 \int_0^\infty \frac{dt}{8t} (16\pi^2 t)^{-5} \eta(it)^{-8} [Z_0^0(it)^4 - Z_0^1(it)^4] \tag{5.120}$$

$$Z_1 = N^2 \int_0^\infty \frac{dt}{8t} (16\pi^2 t)^{-5} \eta(it)^{-8} [Z_1^0(it)^4 + Z_1^1(it)^4]. \tag{5.121}$$

Of course $Z_0 = Z_1$ since the total amplitude Z_{Cy} vanishes. However, just as in the discussion around Eq. (5.108), if we wish the amplitude to be finite for arbitrary *planar* insertions (i.e. insertions on one boundary component only) each amplitude Z_0, Z_1 should be *separately finite* (this statement will be obvious from the analysis below³⁶). Physically, this issue may be understood as follows: divergences in the tree-level closed string amplitude Z_0 arise from infinitely long cylinders,

$$\ell \equiv 1/t \rightarrow \infty, \tag{5.122}$$

which corresponds to the propagation of zero-momentum NS-NS states in the crossed channel. The divergent part of the amplitude Z_0 is then proportional to the square of the disk amplitude with the zero-momentum NS-NS vertex inserted, i.e. to the square of the tadpole amplitude (cf. the analogue discussion for the bosonic string in Sect. 4.9). Using the modular properties in **BOX 5.6**,

³⁵ In Chaps. 6 and 12, we shall re-interpret these b.c. as due to the presence of N spacetime filling D9 branes. See Chaps. 6 and 12 for more details.

³⁶ The analysis will show that the divergence of a planar insertion is the product of some disk amplitude times a tree-level tadpole; the tadpole will vanish if and only if both Z_0, Z_1 are finite.

$$\eta(it) = t^{-1/2} \eta(i/t), \quad Z^\alpha_\beta(it)^4 = Z^\beta_\alpha(i/t)^4 \quad (5.123)$$

we rewrite

$$Z_0 = \frac{N^2}{8(16\pi^2)^5} \int_0^\infty d\ell \, \eta(i\ell)^{-8} [Z^0_0(i\ell)^4 - Z^0_1(i\ell)^4] \quad (5.124)$$

The asymptotics of the function $\eta(\tau)$ as $\ell \rightarrow \infty$ is

$$\eta(i\ell) = e^{-\pi\ell/12} \prod_{n=1}^\infty (1 - e^{-2\pi n\ell}) = e^{-\pi\ell/12} (1 + O(e^{-2\pi\ell})), \quad (5.125)$$

while, using the “abstruse” identity (5.107),

$$\begin{aligned} Z^0_0(i\ell)^4 - Z^0_1(i\ell)^4 &\equiv Z^1_0(i\ell)^4 = \left(\frac{1}{\eta(i\ell)} \cdot \vartheta \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} (0|i\ell) \right)^4 = \\ &= \left(e^{\pi\ell/12} \cdot 2 e^{-\pi\ell/4} (1 + O(e^{-2\pi\ell})) \right)^4 = 16 e^{-2\pi\ell/3} (1 + O(e^{-2\pi\ell})), \end{aligned} \quad (5.126)$$

so that the NS-NS cylindric amplitude

$$Z_0 = \frac{N^2}{8(16\pi^2)^5} \int_0^\infty d\ell \, [16 + O(e^{-2\pi\ell})], \quad (5.127)$$

has a linear divergence which is proportional to the square of a NS-NS tadpole on the disk, analogous to the one in the open bosonic string (cf. discussion in Sect. 4.7). The NS-NS tadpole is given by the disk amplitude with one insertion of the dilaton vertex at zero momentum

$$\left\langle \eta_{\mu\nu} \psi^\mu \tilde{\psi}^\nu e^{-\phi - \tilde{\phi}} \right\rangle_{\text{disk}} \neq 0. \quad (5.128)$$

The R-R amplitude Z_1 has an identical linear divergence which should be interpreted in terms of a R-R tadpole on the disk.

The R-R Tadpole (Solving an Apparent Paradox)

There is no *propagating* R-R 10d field which can be responsible for the above tadpole. Indeed, as discussed in Sect. 3.7.3, in picture $(-\frac{1}{2}, -\frac{1}{2})$ the *propagating* R-R states have vertices proportional to k_μ which vanish in the zero-momentum limit. Equivalently, a non-zero tadpole for a gauge field form,

$$\langle A^{(k)} \rangle \neq 0, \quad (5.129)$$

breaks (besides Lorentz invariance for $k \neq 0$) the spacetime gauge symmetry (i.e. BRST invariance in the world-sheet language) and this is certainly *not* allowed in a consistent theory.

We arrived at an *apparent* paradox, which most string theory textbooks discuss in an incorrect manner. Indeed, consistency with BRST quantization *requires* the R-R tadpole to be the disk amplitude with one R-R-sector *BRST-invariant* operator inserted in the bulk. By CFT state-operator isomorphism, this BRST-invariant operator must correspond to a BRST-invariant *state*. No R-R state visible in light-cone or OCQ will do since their vertices vanish at zero momentum, so—if the states visible in OCQ were the *only* physical states—we would get a contradiction.

Our careful discussion of BRST quantization in Chap. 3 solves this tricky conundrum in a very transparent way; there we showed that *there are more zero-momentum BRST-invariant states than naively expected on the basis of analysis in the light-cone approach (or OCQ)*. The R-R tadpole arises from such subtle BRST-invariant states which *are invisible* in the light-cone gauge or OCQ (quasi-topological modes).

The relevant “subtle” state is easily understood in the light of **BOX 3.3**; in Type IIB the BRST-invariant R-R vertex in the *appropriate*³⁷ $(-\frac{1}{2}, -\frac{3}{2})$ picture has the form

$$\sum_{k \text{ even}} A_{\mu_1 \dots \mu_k}^{(k)}(X) (S \gamma^{\mu_1 \dots \mu_k} \tilde{S}) e^{-\phi/2 - 3\tilde{\phi}/2}, \quad (5.130)$$

and the *even* degree form $A = \sum_k A^{(k)}(x)$ satisfies the Kähler-Dirac equation in $\mathbb{R}^{9,1}$

$$\text{vertex(5.130) is BRST-invariant} \iff (d - \delta)A = 0. \quad (5.131)$$

We know from **BOX 3.3** that A can be chosen to be self-dual, $*A = iA$. Suppose that $A \equiv A^{(10)}$ has *pure degree 10*.³⁸ The BRST condition becomes

$$0 = (d - \delta)A^{(10)} = d(*A^{(10)}) \Rightarrow *A^{(10)} = \text{const}, \quad (5.132)$$

so that taking $A \equiv A^{(10)}$ to be a *constant* 10-form produces a BRST-invariant vertex which is not BRST trivial,³⁹ and hence it must lead to *observable physical effects*. We stress again that this physical vertex does not correspond to any *propagating* 10d° of freedom, since its momentum is frozen to be exactly zero by BRST invariance, see Eq. (5.132).⁴⁰ In other words, while the vertex is physical, we cannot form spacetime wave packets out of it. This is in sharp contrast with the NS-NS vertex responsible

³⁷ We mean *appropriate* for a single bulk R-R insertion on the disk; the left/right-pictures $q_L, q_R \in \frac{1}{2} + \mathbb{Z}$ satisfy $q_L + q_R = -2$. The most canonical solution to these conditions is as in the text.

³⁸ Alternatively we may write $A = A^{(0)} + A^{(10)}$ where $A^{(0)} = -i * A^{(10)}$. The equation $(d - \delta)A = 0$ then yields the two equivalent conditions $d * A^{(10)} = dA^{(0)} = 0$.

³⁹ Since the zero-momentum R vacua are not BRST trivial (see Chap. 3).

⁴⁰ The crucial aspect here is the fact that we emphasized in Chap. 3: BRST cohomology at zero momentum **is not** the $k_\mu \rightarrow 0$ limit of the non-zero momentum BRST cohomology which, in turn, is isomorphic to the light-cone Hilbert space.

for the tadpole in the other sector, Eq. (5.128), which we can “boost” to non-zero momentum while preserving its BRST invariance

$$\eta_{\mu\nu} \psi^\mu \tilde{\psi}^\nu e^{-\phi-\tilde{\phi}} \longrightarrow \eta_{\mu\nu} \psi^\mu \tilde{\psi}^\nu e^{-\phi-\tilde{\phi}} e^{ik \cdot X}, \quad k^2 = 0, \quad (5.133)$$

and hence corresponds to a propagating massless particle; the dilaton Φ . The disk tadpole of the R-R zero-momentum vertex is non-zero

$$\left\langle C^{\alpha\dot{\beta}} S_\alpha \tilde{S}_{\dot{\beta}} e^{-\phi/2-3\tilde{\phi}/2} \right\rangle_{\text{disk}} = \kappa N \neq 0, \quad (5.134)$$

see **BOX 5.8** for more details. The tadpole (5.134) has a factor of N from the trace of the CP labels on the boundary of the disk.

BOX 5.8 - An alternative viewpoint on the RR tadpole

One can easily see that the disk RR tadpole μ does not vanish. A part for an overall normalization constant κ (which we leave to the reader as an **exer**) the tadpole is obtained by inserting in the open disk amplitude the zero-momentum RR vertex in the main text. By the Schwarz reflection principle, we may replace this by the computation on the sphere with only the holomorphic side and the tilted field replaced by holomorphic fields inserted in the symmetric point $\tilde{z}$ of z with respect to the equator. Then the computation of the insertion at zero momentum of the vertex (5.130) in the disk reduces to the evaluation of the chiral amplitude

$$\kappa N \left\langle c(z) S_\alpha(z) e^{-\phi(z)/2} c(1) c(\tilde{z}) S_{\dot{\beta}}(\tilde{z}) e^{-3\phi(\tilde{z})/2} \right\rangle_{S^2}.$$

The asymmetric picture $(-1/2, -3/2)$ is just devised so that in the image method, it produces an operator insertion on the sphere with total picture charge -2 , as required to soak up the two γ zero-modes getting a finite non-zero amplitude. The above amplitude is just $\kappa N C_{\alpha\dot{\beta}}$ and is certainly not zero.

Spacetime Interpretation

Being BRST-invariant in the 2d sense, the zero-momentum R-R vertex corresponds to a *gauge-invariant interaction in 10d target space*. In Fourier analysis, zero momentum means integration over the full $\mathbb{R}^{9,1}$ space, so that the 10d interaction must have the form

$$\mu \int_{\mathbb{R}^{9,1}} A^{(10)}, \quad (5.135)$$

for some non-zero constant μ . This spacetime coupling is indeed both **Diff**⁺-invariant (being topological) as well as gauge invariant

$$\mu \int_{\mathbb{R}^{9,1}} A^{(10)} \equiv \mu \int_{\mathbb{R}^{9,1}} \left(A^{(10)} + d\lambda^{(9)} \right), \quad (5.136)$$

as expected from its 2d BRST invariance. The only possible R-R tadpole is then a 10-form tadpole, and its value is given by the coupling μ in Eq. (5.135). The cylinder amplitude in the R-R channel, Eq. (5.127), shows that the theory of the oriented open superstring necessarily has a non-zero 10-form tadpole μ proportional to N .

The 10d effective action contains $A^{(10)}$ only through the topological coupling (5.135), since the effective Lagrangian $\mathcal{L}_{\text{eff}}$ is local and gauge invariant, and hence—a part for the topological term (5.135)⁴¹— $A^{(10)}$ may enter in $\mathcal{L}_{\text{eff}}$ only through its gauge-invariant field strength

$$F^{(11)} = dA^{(10)} \equiv 0, \quad (5.137)$$

which however is *identically zero* in $\mathbb{R}^{9,1}$. The coupling (5.135) is the only one consistent with gauge invariance, and we see that gauge invariance freezes $A^{(10)}$ to zero momentum, exactly as expected from the world-sheet BRST cohomology.

The equation of motion of $A^{(10)}$ takes the form

$$0 = \frac{\delta \mathcal{S}_{\text{eff}}}{\delta A^{(10)}(x)} \equiv \mu. \quad (5.138)$$

In the oriented open superstring theory, this equation is inconsistent since μ is a non-zero constant $\mu = \kappa N$. It is pretty clear that this tadpole cannot be shifted away; oriented open superstrings are really inconsistent. We already knew that this model is *inconsistent* (cf. Sect. 5.3); we just confirmed our previous conclusion from a different and more “stringy” perspective.

5.6.1 Consistency of $SO(32)$ Type I

Our stringy viewpoint also explains why considering *unoriented* superstrings improves the situation. In perturbative unoriented string theory, the amplitudes are given by the sum over a *larger set* of world-sheet topologies; in addition to the orientable surfaces, we have the non-orientable ones. At one-loop level, besides the torus T^2 and the cylinder Cy , we have the Möbius strip Mö and the Klein bottle Kl . The “new” surfaces contribute to the square of the R-R tadpole⁴²

$$\mu^2 = \mu^2|_{\text{Cy}} + \mu^2|_{\text{Mö}} + \mu^2|_{\text{Kl}}. \quad (5.139)$$

We saw in Eq. (5.127) that

$$\mu^2|_{\text{Cy}} = a_1 N^2 \quad (5.140)$$

⁴¹ In the standard jargon of string theory, one may call (5.135) a “Chern–Simons” coupling.

⁴² The torus does not contribute since the amplitude is finite, see *Remark 5.4*.

for a non-zero constant a_1 which may be read from that equation. The quadratic dependence on N reflects the fact that the cylinder has *two* boundary components; each of them contributes a factor N from the trace over the Chan–Paton d.o.f.

The Möbius strip has one boundary. Going around this boundary we come back with the inverted orientation. From Sect. 3.9, in particular, Eq. (3.311), we see that the trace over the boundary degrees of freedom (i.e. CP labels) produces the factor

$$\begin{aligned} \text{tr}_{\text{CP}}(\Omega) &= \sum_{i,j=1}^N \langle i, j | \Omega | i, j \rangle = \sum_{i,j,i',j'=1}^N \langle i, j | \gamma_{jj'} | j' i' \rangle \gamma_{ii'}^{-1} = \\ &= \delta^{ij'} \delta^{ji'} \gamma_{jj'} \gamma_{ii'}^{-1} = \text{tr}(\gamma^t \gamma^{-1}) = \pm N \end{aligned} \quad (5.141)$$

upper sign for $SO(N)$, lower one for $USp(N)$. Thus

$$\mu^2|_{\text{Mo}} = \pm a_2 N, \quad (5.142)$$

for some constant a_2 . Finally, the Klein bottle has no boundary, and we get

$$\mu^2|_{K^2} = a_3 \quad (5.143)$$

independently of N . Since μ^2 is the square of the sum of the tadpole in the disk and the $\mathbb{RP}^2$ topologies, which have the schematic form $\sqrt{a_1}N$ and $\pm b\sqrt{a_1}$, respectively

$$a_1 N^2 \pm a_2 N + a_3 \equiv \mu^2 = a_1 (N \pm b)^2 \quad \text{so} \quad a_3 = \frac{a_2^2}{4a_1}. \quad (5.144)$$

Hence the tadpole vanishes for the special value of N

$$N = |b| \equiv \frac{|a_2|}{2|a_1|} \equiv \sqrt{\frac{a_3}{a_1}} \quad (5.145)$$

provided it is an integer; the corresponding gauge group is $SO(N)$ for $b > 0$ (resp. $USp(N)$ for $b < 0$). Thus at most *one* gauge group may lead to a consistent Type I theory. As we have already announced, the good group is $SO(32)$. To show this fact, we sketch the computation of the Klein bottle and Möbius strip amplitudes, leaving the details as an **Exercise** for the reader.

Klein Bottle Amplitude

The bosonic and fermionic path integrals on Klein bottle were computed in Sects. 4.9 and 5.4, respectively. There the geometry of this non-orientable surface is also discussed. The dual channel NS-NS amplitude is

$$\frac{1}{2} \sum_{\alpha=0,1} \text{Tr}_{\alpha,\alpha} \left[\Omega (-1)^F e^{-4\pi\tau(L_0 - c/24)} \right]. \quad (5.146)$$

Using the results in Sects. 4.9 and 5.4, we get

$$\begin{aligned}
 \text{dual NS-NS amplitude} &= \int_0^\infty \frac{dt}{8t} (8\pi^2 t)^{-5} \eta(2it)^{-8} [Z_1^0(2it)^4 - Z_1^1(2it)^4] = \\
 &= \frac{1}{8(8\pi^2)^5} \int_0^\infty \frac{dt}{t} \frac{1}{t^5} [(2t)^4 \eta(i/2t)^{-8}] Z_0^1(i/2t)^4 = \quad (5.147) \\
 &= \frac{2^5}{8(8\pi^2)^5} \int_0^\infty d\ell \eta(i\ell)^{-8} Z_0^1(i\ell)^4.
 \end{aligned}$$

Using the “abstruse” identity, we see that (5.147) differs from the corresponding cylinder amplitude Z_0 , Eq. (5.124), only in the overall factor

$$(5.147) = \left(\frac{2^5}{N}\right)^2 Z_0. \quad (5.148)$$

In the language of Eq. (5.145), $a_3/a_1 = 2^{10}N^{-2}$, so that

$$N = 32. \quad (5.149)$$

BOX 5.9 - Ω -twisted bosonic open string partition functions

For a single free real boson X quantized in the strip (with NN b.c.), we have

$$\begin{aligned}
 \text{Tr}[\Omega q^{L_0 - c/24}] &= \pm q^{-1/24} \prod_{n \geq 1} \sum_{N_n=0}^\infty (-q)^{nN_n} = \pm q^{-1/24} \prod_{n \geq 1} (1 - (-q)^n)^{-1} = \\
 &= \pm q^{-1/24} \prod_{n \geq 1} (1 + q^{2n-1})^{-1} (1 - q^{2n})^{-1}
 \end{aligned}$$

where the $\pm$ sign depends on the action of Ω on the vacuum: $\Omega|0\rangle = \pm|0\rangle$. From Eq. (5.71), we have

$$\prod_{m=1}^\infty (1 + q^{m-1/2}) = \left(q^{1/24} Z_0^0(q)\right)^{1/2},$$

so

$$\text{Tr}[\Omega q^{L_0 - c/24}] = \pm \frac{1}{\eta(q^2) \sqrt{Z_0^0(q^2)}}.$$

To distinguish $SO(32)$ from $USp(32)$, we need the Möbius strip amplitude.

Möbius Strip Amplitude

The relevant Bose/Fermi path integrals were computed in Sects. 4.9, 5.4. The Bose one is also reviewed in BOX 5.9. We focus on the R-R exchange⁴³

$$\begin{aligned} Z_{\text{Mo,RR}} &= -\frac{1}{2} \sum_{\beta=0,1} \text{Tr}_R [\Omega (-1)^{\beta F} q^{(L_0-c/24)}] = \\ &= -\text{tr}_{\text{CP}}(\Omega) \int_0^\infty \frac{dt}{4t} (16\pi^2 t)^{-5} \frac{Z_1^0(2it)^4 Z_0^1(2it)^4}{\eta(2it)^8 Z_0^0(2it)^4}. \end{aligned} \quad (5.150)$$

The CP factor $\text{tr}_{\text{CP}}(\Omega)$ is computed in Eq. (5.141).

Exploiting the modular properties of the functions, we rewrite Eq. (5.150) as

$$Z_{\text{Mo,RR}} = \mp \frac{2 \cdot 2^5 N}{8(16\pi^2)^5} \int_0^\infty d\ell \frac{Z_1^0(i\ell)^4 Z_0^1(i\ell)^4}{\eta(i\ell)^8 Z_0^0(i\ell)^4}. \quad (5.151)$$

Using Eq. (5.126) and the large ℓ asymptotics

$$Z_1^0(i\ell)^4, Z_0^0(i\ell)^4 = e^{-\pi\ell/4} (1 + O(e^{-2\pi\ell})) \quad (5.152)$$

we get

$$Z_{\text{Mo,RR}} = \mp \frac{2 \cdot 2^5 N}{8(16\pi^2)^5} \int_0^\infty d\ell (16 + O(e^{-2\pi\ell})). \quad (5.153)$$

Then the total divergence in the R-R exchange is

$$\begin{aligned} Z_{\text{Cy,RR}} + Z_{\text{Mo,RR}} + Z_{\text{K}^2,\text{RR}} &= \\ &= (N^2 \mp 2 \cdot 2^5 N + 2^{10}) \frac{1}{8(16\pi^2)^5} \int_0^\infty d\ell (16 + O(e^{-2\pi\ell})). \end{aligned} \quad (5.154)$$

We see that the R-R tadpole vanishes only for $G = SO(32)$.

The NS-NS divergence is the negative of the R-R one, so it also vanishes for $G = SO(32)$. Note that the coefficient is a perfect square $(N \mp 32)^2$, as it should be, since by unitarity it should coincide with the square of the sum of the disk and $\mathbb{RP}^2$ NS-NS tadpoles.

⁴³ The overall minus sign reflects (from one-loop open channel viewpoint) that R sector open string states are fermions.

References

1. Farb, B., Margalit, D.: A primer on Mapping Class Groups. Princeton University Press (2012)
2. Spanier, E.H.: Algebraic Topology. Springer, Berlin (1966)
3. Farkas, H.M., Kra, I.: Riemann Surfaces. Graduate Texts in Mathematics, vol. 71. Springer, Berlin (1992)
4. Griffiths, P., Harris, J.: Principles of Algebraic Geometry. Wiley (1978)
5. Carlson, J., Müller-Stach, S., Peters, C.: Period Mappings and Period Domains, vol. 168, 2nd edn. Cambridge Studies in Advanced Mathematics. Cambridge University Press (2017)
6. Cecotti, S.: Supersymmetric Field Theories. Cambridge University Press, Geometric Structures and Dualities (2015)
7. Helgason, S.: Differential Geometry, Lie Groups and Symmetric Spaces. Academic Press (1978)
8. Fay, J.D.: Theta Functions on Riemann Surfaces. Lecture Notes in Mathematics, vol. 352. Springer, Berlin (1973)
9. Dijkgraaf, R., Verlinde, E., Verlinde, H.: $c = 1$ conformal field theories on Riemann surfaces. Comm. Math. Phys. **115**, 649–690 (1988)
10. Verlinde, E., Verlinde, H.: Multiloop calculations in covariant superstring theory. Phys. Lett. B **192**, 95–102 (1987)
11. Borel, A.: Introduction aux Groupes Arithmétiques. Hermann (1969)
12. Serre, J.-P.: Bounds for the orders of the finite subgroups of $G(k)$. In: Geck, M., Testerman, D., Thévenaz, J. (eds.) Group Representation Theory. EPFL Press (2006)
13. Deser, S., Zumino, B.: Consistent supergravity. Phys. Lett. B **62**, 335 (1976)
14. Coleman, S., Weinberg, E.: Radiative corrections as the origin of spontaneous symmetry breaking. Phys. Rev. D **7**, 1888 (1973)

Chapter 6

Bosonic String: T -Duality & D -Branes



Abstract In this textbook, the main emphasis is on *supersymmetric* string theories. We use the bosonic string merely as a kindergarten laboratory to introduce ideas and techniques in the simplest possible context. The extension of these results to the supersymmetric situation is then conceptually natural, despite many new technicalities and subtleties. Following our didactical strategy, in this chapter, we discuss in the bosonic set-up: (i) T -duality, (ii) Buscher rules, (iii) Narain compactifications, (iv) the effects of Wilson lines, (v) the stringy Higgs mechanism, (vi) D -branes, and (vii) orientifolds. In the process: (a) we study the CFT of compact scalars with emphasis on the chiral ones, (b) we outline the relations with lattice theory and arithmetic quotients, (c) we give a quick introduction to orbifolds, and (d) prove general re-fermionization identities (a.k.a. Riemann relations).

6.1 Toroidal Compactifications in Field Theory

In General Relativity, the geometry of spacetime is dynamical. Hence we may (and should) consider spacetimes of any geometry/topology. The same holds for the bosonic string since it contains Einstein's gravity. For the sake of comparison, we start by looking at toroidal compactifications in field theory and then study their new surprising features in string theory.

Kaluza–Klein Geometry

Consider a spacetime of dimension $D \equiv d + 1$ diffeomorphic to

$$\mathbb{R}^{d-1,1} \times S^1. \quad (6.1)$$

We write $y \equiv x^d$ for the periodic coordinate along S^1

$$y \sim y + 2\pi R, \quad (6.2)$$

and x^μ ($\mu = 0, 1, \dots, d-1$) for the coordinates in the non-compact factor $\mathbb{R}^{d-1,1}$. We write the background metric in the form¹

$$ds^2 = G_{MN} dx^M dx^N = G_{\mu\nu} dx^\mu dx^\nu + e^{2\sigma} (dy + A_\mu dx^\mu)^2 \quad (6.3)$$

$G_{\mu\nu} \equiv G_{\mu\nu} - e^{2\sigma} A_\mu A_\nu$ is the d -dimensional metric. For the moment, we allow the fields $G_{\mu\nu}$, A_μ , and σ to depend only on the non-compact directions x^μ . With this restriction, $\partial/\partial y$ is a nowhere vanishing Killing vector. In fact, Eq. (6.3) is the most general metric invariant under translations² in y .

The form (6.3) of the metric is preserved by d -dimensional diffeomorphisms $x^\mu \rightarrow x'^\mu(x^\nu)$ as well as by reparametrizations of the form

$$y' = y + \lambda(x^\mu) \quad (6.4)$$

under which

$$A_\mu \rightarrow A'_\mu = A_\mu - \partial_\mu \lambda. \quad (6.5)$$

Thus the d -dimensional Abelian gauge transformations (6.5) arise from the higher dimensional diffeomorphisms: this is the *Kaluza–Klein (KK) mechanism* [1–3].

Modern Viewpoint: Principal $U(1)$ Bundles There is a more intrinsic viewpoint on the KK geometry, which covers topologies more general and physically interesting than the “trivial” product (6.1). Consider a (pseudo-)Riemannian manifold M which has a nowhere vanishing space-like Killing vector K whose orbits are all closed, hence diffeomorphic to $U(1) \equiv S^1$. The action of $U(1)$ on M is free, so we have a quotient manifold $B = M/U(1)$, the *orbit space*. The canonical projection

$$\pi : M \rightarrow B \quad (6.6)$$

which associates to a point its $U(1)$ orbit is a $U(1)$ -principal bundle. By definition, the group $\mathcal{G}$ of $U(1)$ gauge transformations of M is the automorphisms group of the principal bundle $\pi : M \rightarrow B$, i.e. the group of fiber-preserving $U(1)$ -equivariant diffeomorphisms of M , see, e.g. Chap. 7 of [4].

Proposition 6.1 *All metrics on the total space M of the $U(1)$ -principal bundle $\pi : M \rightarrow B$ which are gauge-invariant (i.e. their form is preserved by $\mathcal{G}$ while “rigid” $U(1)$ acts by isometries) have the form*

$$ds^2 = ds_B^2 - e^{2\sigma} \theta^2 \quad (6.7)$$

¹ **Convention:** D -dimensional geometric quantities are written boldface, d -dimensional ones in normal type. All metrics on $\mathbb{R}^{d-1,1} \times S^1$ can be written as in (6.3).

² That is, all metrics on a smooth manifold M which have a nowhere vanishing Killing vector K , whose orbits are all closed, is locally isometric to a metric of the form (6.3) with $G_{\mu\nu}$, A_μ and σ independent of y . For the proof of a more general statement, see **Proposition 6.1** below.

for some metric ds_B^2 on the base B , connection form θ on the principal bundle,³ and function σ on B . Conversely all metrics (6.7) are gauge-invariant in the above sense.

Proof The Lie group $U(1)$ acts freely on ds^2 by isometries, so there is a *nowhere vanishing* Killing vector $K^i \partial_i$. The metric defines a dual form $\kappa \equiv K_i dx^i$ such that $\mathcal{L}_K \kappa = 0$. Then κ and $\iota_K \kappa$ are invariant by translation along the $U(1)$ fibers. Set

$$e^{2\sigma} \equiv \iota_K \kappa, \quad -i\theta \equiv e^{-2\sigma} \kappa. \quad (6.8)$$

$e^{2\sigma} \equiv K^i K_i$ is a positive function on B , while θ is $U(1)$ invariant one-form on M which acts as the identity⁴ on the tangent space to the fiber, i.e. θ is a $U(1)$ -principal connection form [5, 6]. $\square$

On the base B , we have the curvature 2-form

$$F \stackrel{\text{def}}{=} d\theta \equiv dA \in \Omega^2(B), \quad dF = 0, \quad (6.9)$$

whose cohomology class is proportional to the first Chern class of the bundle

$$c_1(M) = \left[\frac{F}{2\pi i} \right] \in H^2(B, \mathbb{Z}). \quad (6.10)$$

The manifold M is topologically a product as in (6.1) iff $\pi : M \rightarrow B$ is a *trivial* principal bundle, i.e. $c_1 \equiv 0$, namely iff the flux $\int_S F$ of the magnetic field F through any closed surface $S \subset B$ vanishes, that is, if no $U(1)$ magnetic monopole⁵ is present. When our geometry has $c_1 \neq 0$, we say that it is a *KK monopole*.

The metric (6.3) has the form (6.7) with the standard identification

$$-i\theta = dy + A_\mu dx^\mu, \quad (6.11)$$

where y is the coordinate along the fiber, while $c_1(M) = 0$ and the bundle is a smooth product, Eq. (6.1).

Example: the Hopf fibration as a KK monopole

A celebrated KK geometry with a topologically non-trivial gauge principal bundle is the *Hopf fibration* $\pi : S^3 \rightarrow S^2$ [4]. We parametrize the unit 3-sphere $S^3 \subset \mathbb{R}^4 \simeq \mathbb{C}^2$, whose equation is $|z_1|^2 + |z_2|^2 = 1$, with angles θ, ϕ, ψ as

$$z_1 = e^{i(\psi+\phi)/2} \cos(\theta/2), \quad z_2 = e^{i(\psi-\phi)/2} \sin(\theta/2). \quad (6.12)$$

The round metric on S^3 is

$$ds^2 = \left(d\theta^2 + \sin^2 \theta d\phi^2 \right) + \left(d\psi - \cos(\theta) d\phi \right)^2, \quad (6.13)$$

the first term is the round metric on the base S^2 and the second one is 4 times the square of the connection form $i\theta = \frac{1}{2}(d\psi - \cos(\theta) d\phi)$. This Abelian gauge field describes the constant

³ For the connection forms on principal bundles, see [5, 6].

⁴ We identify $i\mathbb{R}$ with the Lie algebra of $U(1)$.

⁵ We use the term “monopole” in a general sense; in 4d, it reduces to the usual notion.

magnetic field on the unit sphere $S^2 \subset \mathbb{R}^3$ induced by a magnetic monopole of charge 1 at the origin of $\mathbb{R}^3$

$$F = \frac{i}{2} \sin(\theta) d\theta \wedge d\phi. \quad (6.14)$$

An extra bonus of the principal bundle viewpoint is that it allows for a straightforward generalization of the KK mechanism to produce gauge theories in dimension $d = D - \dim G$ for any compact *non-Abelian* gauge group G with effective lower-dimensional gauge field A in an arbitrary topological class, see **BOX 6.1**.

Kaluza–Klein (KK) Modes

We consider the KK geometry (6.3) where $G_{\mu\nu}$, A_μ and σ are arbitrary but y independent. We write $D \equiv d + 1$ for the dimension of the total space. In this background geometry, we consider a D -dimensional free massless scalar $\phi(x^M)$ whose e.o.m. are

$$\square\phi \equiv G^{MN} \nabla_M \nabla_N \phi = 0, \quad (6.15)$$

with ∇_M (resp. ∇_μ) the Levi-Civita connection in D (resp. d) dimensions.

Exercise 6.1 Show that for the D -dimensional metric under consideration

$$\square = G^{\mu\nu} D_\mu D_\nu + e^{-2\sigma} \partial_y^2 + G^{\mu\nu} \partial_\mu \sigma D_\nu \quad \text{where} \quad D_\mu \equiv \nabla_\mu - A_\mu \partial_y. \quad (6.16)$$

We expand the field $\phi(x^M)$ in Fourier modes along S^1

$$\phi(x^M) = \sum_{n \in \mathbb{Z}} \phi_n(x^\mu) e^{iny/R}. \quad (6.17)$$

The D -dimensional e.o.m. (6.15) become

$$-D_\mu D^\mu \phi_n + \frac{n^2}{R^2} e^{-2\sigma} \phi_n - \partial^\mu \sigma D_\mu \phi_n = 0 \quad (6.18)$$

where D_μ is the gravitational and *gauge* covariant derivative

$$D_\mu \phi_n = \left(\nabla_\mu - i \frac{n}{R} A_\mu \right) \phi_n. \quad (6.19)$$

The meaning of these equations is clear: from a d -dimensional perspective, the n th Fourier mode $\phi_n(x^\mu)$ is a field of electric charge n coupled to the KK photon A_μ . We are mainly interested in small fluctuations around a “vacuum” configuration where the KK scalar σ is constant. The n th KK mode propagates in such a background as a d -dimensional charged particle of mass

$$m_n^2 = \frac{n^2}{R^2} e^{-2\langle\sigma\rangle}. \quad (6.20)$$

A single scalar field in D dimensions yields an infinite tower of scalar fields in d -dimensions: the n th field has KK charge n and mass proportional to $|n|$. We call this infinite set of fields/particles a *KK tower*.

Above we considered a D -dimensional *scalar* field. It is obvious from the argument that the conclusions hold for D -dimensional fields of any spin. *Mutatis mutandis* we get a tower of lower-dimensional fields—all of the same spin—of KK charge $n \in \mathbb{Z}$ and mass growing linearly with n .

BOX 6.1 - Non-Abelian Kaluza-Klein geometries

The KK construction may be generalized to an arbitrary compact Lie group G .

For definiteness we take G to be *simple*, the general case being then obvious. If $\pi: M \rightarrow B$ is a G -principal bundle and θ a G -connection form [5, 6], the G -invariant metrics on M are

$$ds^2 = ds_B^2 - e^{2\sigma} \text{tr}(\theta^2)$$

where $\text{tr}(\cdot)$ stands for the Killing form on $\mathfrak{g} = \mathfrak{Lie}(G)$. The non-Abelian gauge transformations on the base B act on the total space M by (smooth) automorphisms which preserve the form of the metric, and M has $\dim G$ linear-independent Killing vectors generating the Lie algebra $\mathfrak{g}$. The topological classification of the non-Abelian KK geometries is based on the *characteristic classes* [4, 7, 8] of the associated principal bundle $\pi: M \rightarrow B$.

We stress that (from the D -dimensional perspective) the momentum in the compact direction (i.e. the eigenvalue of the differential operator $-i\partial_y$) gets quantized

$$p_d = \frac{n}{R}, \quad n \in \mathbb{Z}. \quad (6.21)$$

The second Eq. (6.16) shows that from the d -dimensional perspective the compact momentum is just the KK electric charge, which is quantized in integral units because the gauge group $U(1)$ is compact.

At energies $\ll R^{-1}$, all modes ϕ_n decouple except the constant mode $\phi_0(x^\mu)$, and the physics becomes purely d -dimensional. At energies larger than R^{-1} , the tower of Kaluza–Klein states ϕ_n is visible and the physics looks really D -dimensional.⁶

We restate our results in a fancier language which makes sense even for non-trivial KK monopoles. Recall that for each $U(1)$ -module⁷ V there is a vector bundle

⁶ This shows that going to the IR we lose information about the massive modes, so *in field theory* the RG flow is “irreversible”. We already know that often in string theory, the UV limit may be reinterpreted (more conveniently) as an IR limit; this implies that in the IR limit, we have the *same* amount of information as in the UV limit. In particular, a string theory in the IR limit should also behave as a *full-fledged* D -dimensional theory. We shall see below that this is indeed the case.

⁷ If G is a Lie group, by a G -module we mean the representation space of a linear G representation.

$O(V) \rightarrow B$, with typical fiber V and structure group $U(1)$, which is associated to our $U(1)$ -principal bundle $\pi: M \rightarrow B$:

$$O(V) \stackrel{\text{def}}{=} M \otimes V / \{(mu^{-1}, u \cdot v) \sim (m, v), u \in U(1)\}. \quad (6.22)$$

Let $\mathcal{L} \equiv O(\chi_1)$ be the line bundle associated to fundamental character of $U(1)$ (the embedding $\chi_1: U(1) \hookrightarrow \mathbb{C}^\times$). The above discussion may be condensed in the equality⁸

$$L^2(M) = \overline{\bigoplus_{n \in \mathbb{Z}} C_c^\infty(B, \mathcal{L}^n)}. \quad (6.23)$$

The closure of the n th summand in (6.23) is seen as the Hilbert space of wave functions in B for a particle of charge n . In view of the Peters–Weyl theorem [9], a similar equality holds for KK geometries with any compact gauge group.

Effective Action for Massless Fields

We consider the *field theoretic* d -dimensional low-energy effective action obtained by compactifying the $D = d + 1$ dimensional effective action of Sect. 1.8.1, i.e. obtained by restricting the several D -dimensional fields to their constant modes in the compactified direction. The D -dimensional metric G_{MN} is taken of the form (6.3).

Exercise 6.2 We write G_{MN} for the D -dimensional metric (6.3) and R for its scalar curvature; $G_{\mu\nu}$ and R stand for the corresponding $d \equiv D - 1$ dimensional quantities. With $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, show that

$$\sqrt{-G} R = \sqrt{-G} \left[e^\sigma R - 2 \nabla_\mu \nabla^\mu e^\sigma - \frac{1}{4} e^{3\sigma} F_{\mu\nu} F^{\mu\nu} \right]. \quad (6.24)$$

After compactification on a circle S^1 the graviton-dilaton effective action (1.164) (with $H_\mu \nu \rho \equiv 0$) becomes

$$\begin{aligned} & \frac{1}{2\kappa_0^2} \int d^D x \sqrt{-G} e^{-2\Phi} \left(R + 4 \nabla_M \Phi \nabla^M \Phi \right) = \\ & = \frac{\pi R}{\kappa_0^2} \int d^d x \sqrt{-G} e^{-2\Phi+\sigma} \left(R - 4 \partial_\mu \Phi \partial^\mu \sigma + 4 \partial_\mu \Phi \partial^\mu \Phi - \frac{1}{4} e^{2\sigma} F_{\mu\nu} F^{\mu\nu} \right) = \\ & = \frac{\pi R}{\kappa_0^2} \int d^d x \sqrt{-G} e^{-2\Phi} \left(R - \partial_\mu \sigma \partial^\mu \sigma + 4 \partial_\mu \Phi \partial^\mu \Phi - \frac{1}{4} e^{2\sigma} F_{\mu\nu} F^{\mu\nu} \right) \end{aligned} \quad (6.25)$$

where

$$\Phi = \Phi - \frac{1}{2} \sigma \quad (6.26)$$

⁸ If $V \rightarrow B$ is a smooth vector bundle, we write $C_c^\infty(B, V)$ for the vector $\mathbb{C}$ -space of smooth sections with compact support in the base B . The over-bar stands for the Hilbert space closure.

is the effective d -dimensional dilaton. The apparent wrong sign of the dilaton kinetic term is compensated by the mixing with the graviton's kinetic terms.

Exercise 6.3 Go to the Einstein frame and check the sign of the dilaton kinetic term.

The equations of motion do not fix the radius of compactification or Φ ; indeed σ and Φ have no potential, and the flat metric is a solution to the equation of motion for any periodicity of the compactified coordinate y .

Antisymmetric Tensor Field

The bosonic string low-energy effective Lagrangian (1.164) also contains the antisymmetric tensor B_{MN} . The field B_{MN} also leads to a gauge vector in $d = D - 1$ dimensions. Indeed, from B_{MN} , we get in $d \equiv D - 1$ dimension a two-form field $B_{\mu\nu} \equiv B_{\mu\nu}$ and one vector $B_\mu \equiv B_{\mu y}$; in the same way, the 1-form gauge parameter ξ_M splits in d -dimension in a vector parameter $\xi_\mu(x^v) \equiv \xi_\mu$ and a scalar one $\phi(x^v) \equiv \xi_y$ with

$$B_\mu \rightarrow B_\mu + \partial_\mu \phi(x^v) \quad (6.27)$$

and gauge-invariant field strength 2-form on B

$$H_{\mu\nu}^{(2)} = \partial_\mu B_\nu - \partial_\nu B_\mu. \quad (6.28)$$

Improved Field Strength The gauge-invariant 3-form field strength $\tilde{H}^{(3)}$ of the d -dimensional 2-form field $B^{(2)} \equiv \frac{1}{2} B_{\mu\nu} dx^\mu \wedge dx^\nu$ is a bit subtle because of a mixing between the $U(1)$ gauge symmetry (6.5) and the 2-form gauge symmetry

$$B^{(2)} \rightarrow B^{(2)} + d\xi^{(1)}, \quad \xi^{(1)} \equiv \left[\begin{array}{l} \text{1-form in the } d \text{ non-} \\ \text{compact dimensions} \end{array} \right] \quad (6.29)$$

The simplest way to get the correct expression for $\tilde{H}^{(3)}$ is to compute the field strength in D -dimensions. Let E_a^M be the inverse vielbein in D -dimensions and e_a^μ the one in d -dimensions, where $a = 0, 1, \dots, d - 1$ is the “flat” index. One has

$$\begin{aligned} E_a^M E_b^N E_c^P H_{MNP} &= e_a^\mu e_b^\nu e_c^\rho \left(H_{\mu\nu\rho} - A_\mu H_{\nu\rho}^{(2)} - A_\nu H_{\rho\mu}^{(2)} - A_\rho H_{\mu\nu}^{(2)} \right) \equiv \\ &\equiv e_a^\mu e_b^\nu e_c^\rho \left(\partial_\mu B_{\nu\rho} - A_\mu H_{\nu\rho}^{(2)} + \text{cyclic permutations} \right) \end{aligned} \quad (6.30)$$

This shows that the gauge-invariant field strength is not just dB but

$$\tilde{H}_{\mu\nu\rho} = \partial_\mu B_{\nu\rho} - A_\mu H_{\nu\rho}^{(2)} + \text{cyclic permutations}, \quad (6.31)$$

$$\text{that is, in form notation:} \quad \tilde{H}^{(3)} = dB^{(2)} - A^{(1)} \wedge H^{(2)}. \quad (6.32)$$

Such modified field strengths are called *improved*.⁹ The improved field strength $\tilde{H}^{(3)}$ is invariant (besides for the d -dimensional 2-form gauge symmetry (6.29)) under the KK Abelian gauge transformation (6.5); indeed the B -field transforms as

$$B^{(2)} \rightarrow B^{(2)} - \lambda H^{(2)}, \quad (6.33)$$

see **BOX 6.2**. The gauge-invariant field strength 3-form $\tilde{H}$ satisfies the Bianchi identity

$$d\tilde{H}^{(3)} = -F^{(2)} \wedge H^{(2)}. \quad (6.34)$$

After the compactification on S^1 , the B -field action becomes

BOX 6.2 - Abelian gauge transformation of $B_{\mu\nu}$, Eq. (6.33)

The D -dimensional 2-form

$$B_{MN} dx^M \wedge dx^N = B_{\mu\nu} dx^\mu \wedge dx^\nu + 2 B_{\mu d} dx^\mu \wedge dx^d$$

transforms under the diffeomorphism (6.4) as

$$\begin{aligned} & B_{\mu\nu} dx^\mu \wedge dx^\nu + 2 B_{\mu d} dx^\mu \wedge dx^d - 2 B_{\mu d} dx^\mu \wedge d\lambda = \\ &= \left(B_{\mu\nu} + \partial_\mu \lambda B_{\nu d} - \partial_\nu \lambda B_{\mu d} \right) dx^\mu \wedge dx^\nu + \dots = \\ &= \left(B_{\mu\nu} + \partial_\mu (\lambda B_{\nu d}) - \partial_\nu (\lambda B_{\mu d}) - \lambda (\partial_\mu B_{\nu d} - \partial_\nu B_{\mu d}) \right) dx^\mu \wedge dx^\nu + \dots \end{aligned}$$

so that

$$B_{\mu\nu} \rightarrow B_{\mu\nu} - \lambda H_{\mu\nu d} + 2\text{-form gauge transformation}$$

$$\begin{aligned} & -\frac{1}{24\kappa_0^2} \int d^D x \sqrt{-G} e^{-2\Phi} H_{MNL} H^{MNL} = \\ &= -\frac{\pi R}{12\kappa_0^2} \int d^d x \sqrt{-G} e^{-2\Phi} \left(\tilde{H}_{\mu\nu\rho} \tilde{H}^{\mu\nu\rho} + 3e^{-2\sigma} H_{\mu\nu}^{(2)} H^{(2)\mu\nu} \right). \quad (6.35) \end{aligned}$$

⁹ As a matter of **notation**, in these notes, a *tilde* over the symbol of a field strength, $\tilde{F}$ or $\tilde{H}$, will always mean that the field strength is an improved one, that is, its expression contains additional non-linear terms besides the ones in the usual expressions $F = dA$ or $H = dB$.

6.2 2d CFT of a Compact Scalar

We now focus on the bosonic string moving in the product space $\mathbb{R}^{d-1,1} \times S^1$ with locally flat metric. The world-sheet action decomposes into d non-compact free scalars, already studied in Sect. 2.4, and one compact scalar for the coordinate along S^1 . In this section, we study the 2d CFT of a *free compact* scalar field $X(z, \bar{z})$ which takes value in a circle of radius R , i.e. the scalar field $X(z, \bar{z})$ gets periodically identified¹⁰

$$X \sim X + 2\pi R. \quad (6.36)$$

The action reads (as in the non-compact model)

$$\frac{1}{2\pi\alpha'} \int d^2z \, \partial X \bar{\partial} X. \quad (6.37)$$

The equations of motion, OPEs, and energy–momentum tensor $T(z)$ have the same form as in the non-compact theory, so the theory is still conformally invariant.

The periodicity (6.36) has two effects. **(1)** String states must be single valued in

$$S^1 \equiv \mathbb{R}/2\pi R \mathbb{Z}, \quad (6.38)$$

thus the operator $\exp(2\pi i R P)$ which translates strings around the periodic direction must leave all states invariant. The center of mass momentum k is then quantized

$$k = \frac{n}{R}, \quad n \in \mathbb{Z}. \quad (6.39)$$

This effect was already present in field theory; cf. Eq. (6.21).

(2) A closed string may *wind* around S^1 , i.e. satisfy the periodic b.c.

$$X(\sigma + 2\pi) = X(\sigma) + 2\pi R w, \quad w \in \mathbb{Z}. \quad (6.40)$$

The integer w is a new quantum number known as the *winding number*. From the viewpoint of the world-sheet, QFT strings of non-zero winding number are *topological solitons*, i.e. topologically non-trivial solutions of the e.o.m. The non-triviality of their topology is measured by the topological charge

$$w = \oint \frac{dX}{2\pi R} \in \mathbb{Z} \quad (6.41)$$

called *winding number*. A consistent string theory moving in $\mathbb{R}^{d-1,1} \times S^1$ must include the winding states; a $w = 0$ string can evolve into a $w = +1$ and a $w = -1$

¹⁰ In a different language: we gauge the group of translations by multiples of $2\pi R$ of the non-compact scalar with target space $\mathbb{R}$ and action (6.37).

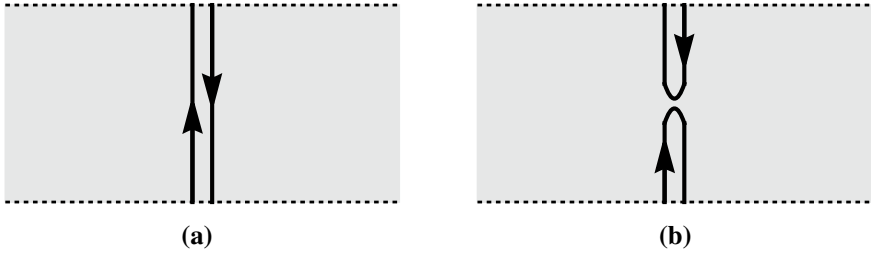


Fig. 6.1 **a** Two strings wrapped on a cylinder with winding $w = +1$ and $w = -1$, respectively. Dashed horizontal lines are identified **b** A single closed string with zero winding number, $w = 0$, obtained by recombination of the two strings in Fig. 6.1a

pair by the splitting–joining process, see Fig. 6.1a, b.¹¹ In *closed oriented* string theory, the total winding number is conserved in all processes, since w_{tot} is a topological charge, see **BOX 6.3**.

Mode Expansions

Consider the Laurent mode expansions

$$\partial X(z) = -i \left(\frac{\alpha'}{2} \right)^{1/2} \sum_{m=-\infty}^{+\infty} \frac{\alpha_m}{z^{m+1}}, \quad \bar{\partial} X(z) = -i \left(\frac{\alpha'}{2} \right)^{1/2} \sum_{m=-\infty}^{+\infty} \frac{\tilde{\alpha}_m}{\bar{z}^{m+1}}. \quad (6.42)$$

The change in the coordinate X as we go around the closed string is

$$2\pi R w = \oint (dz \partial X + d\bar{z} \bar{\partial} X) = 2\pi (\alpha'/2)^{1/2} (\alpha_0 - \tilde{\alpha}_0), \quad (6.43)$$

while the total Noether momentum is

$$P = \frac{1}{2\pi\alpha'} \oint (dz \partial X - d\bar{z} \bar{\partial} X) = (2\alpha')^{-1/2} (\alpha_0 + \tilde{\alpha}_0). \quad (6.44)$$

For a *non-compact* field, this gives the zero-modes in terms of the momentum p

$$\text{non-compact scalar} \quad \alpha_0 = \tilde{\alpha}_0 = (\alpha'/2)^{1/2} p, \quad (6.45)$$

while for a *compact* scalar, we have

$$p_L \equiv (2/\alpha')^{1/2} \alpha_0 = \frac{n}{R} + \frac{wR}{\alpha'}, \quad p_R \equiv (2/\alpha')^{1/2} \tilde{\alpha}_0 = \frac{n}{R} - \frac{wR}{\alpha'}, \quad (6.46)$$

¹¹ E.g. consider the field configuration on $\mathbb{P}^1 \setminus (D_0 \cup D_1 \cup D_\infty)$ where D_z stands for a small disk centered at $z \in \mathbb{P}^1$: $X(z, \bar{z}) = iR(\log(\bar{z} - 1) - \log(z - 1))$ which has an initial state on $-\partial D_0$ with $w = 0$ ending in two final states; one with $w = +1$ on ∂D_1 and one with $w = -1$ on ∂D_∞ .

with $n, w \in \mathbb{Z}$. The Virasoro generators now are

$$L_0 = \frac{\alpha' p_L^2}{4} + \sum_{n \geq 1} \alpha_{-n} \alpha_n, \quad \tilde{L}_0 = \frac{\alpha' p_R^2}{4} + \sum_{n \geq 1} \tilde{\alpha}_{-n} \tilde{\alpha}_n. \quad (6.47)$$

The Partition Function: Operator Approach

The partition function of the compact scalar X is the product of the zero-mode and oscillator contributions. The oscillator sum is the same one as in the non-compact case, producing the usual factor $|\eta(\tau)|^{-2}$, while in the zero-mode sector, the integration over the continuous momentum p is replaced by a discrete sum over the integers n and w . We get

$$\begin{aligned} Z(R, \tau)_{S^1} &\stackrel{\text{def}}{=} (q\bar{q})^{-1/24} \text{Tr} \left(q^{L_0} \bar{q}^{\tilde{L}_0} \right) = |\eta(\tau)|^{-2} \sum_{n, w = -\infty}^{+\infty} q^{\alpha' p_L^2/4} \bar{q}^{\alpha' p_R^2/4} = \\ &= |\eta(\tau)|^{-2} \sum_{n, w = -\infty}^{+\infty} \exp \left[-\pi \tau_2 \left(\frac{\alpha' n^2}{R^2} + \frac{w^2 R^2}{\alpha'} \right) + 2\pi i \tau_1 n w \right]. \end{aligned} \quad (6.48)$$

Modular Invariance Modular invariance of (6.48) is not manifest, but it can be easily checked using the (special case of the) *Poisson resummation formula*

$$\sum_{n=-\infty}^{+\infty} e^{-\pi a n^2 + 2\pi i b n} = a^{-1/2} \sum_{m=-\infty}^{+\infty} e^{-\pi (m-b)^2/a} \quad (6.49)$$

The partition function of the compact scalar of period $2\pi R$ then becomes see **BOX 6.4** for details, proofs, and generalizations.

$$2\pi R Z_X(\tau) \sum_{m, w = -\infty}^{+\infty} \exp \left(-\frac{\pi R^2 |m - w\tau|^2}{\alpha' \tau_2} \right) \quad (6.50)$$

where $Z_X(\tau)$ is the *modular-invariant* partition function of the non-compact theory

$$Z_X(\tau) = (4\pi^2 \alpha' \tau_2)^{-1/2} |\eta(\tau)|^{-2}. \quad (6.51)$$

The sum in Eq. (6.50) is invariant under $\tau \rightarrow \tau + 1$ since it may be undone by redefining the summation variable $m \rightarrow m + w$. Likewise, it is also invariant under $\tau \rightarrow -1/\tau$ together with $m \rightarrow -w$ and $w \rightarrow m$. A more intrinsic derivation of modular invariance of (6.50) (in a form that we shall use repeatedly in this book) is presented in **BOX 6.5** by relating the expression to the $SL(2, \mathbb{Z})$ *Hodge norm* in the sense of theory of *variations of Hodge structure* (VHS) [11–13].

BOX 6.3 - Conservation of winding number

We focus on closed oriented strings propagating in $\mathbb{R}^{25} \times S^1$ with X the periodic coordinate. The winding number of a closed string is

$$w = \oint \frac{dX}{2\pi R}$$

where the integral is along the string. Consider a g loops string amplitude with n_{in} (resp. n_{out}) incoming (resp. outgoing) strings. The amplitude is a path integral over a genus g surface Σ with an initial (resp. final) boundary B_{in} (resp. B_{out}) consisting of n_{in} (resp. n_{out}) disjoint circles

$$B_{\text{in}} = \coprod_{i=1}^{n_{\text{in}}} S_{\text{in},i}^1, \quad B_{\text{out}} = \coprod_{a=1}^{n_{\text{out}}} S_{\text{out},a}^1.$$

The total initial and final winding number is

$$w_{\text{in}} = \int_{B_{\text{in}}} \frac{dX}{2\pi R} \equiv \sum_{i=1}^{n_{\text{in}}} \oint_{S_{\text{in},i}^1} \frac{dX}{2\pi R}, \quad w_{\text{out}} = \int_{B_{\text{out}}} \frac{dX}{2\pi R}$$

Since

$$\partial \Sigma = B_{\text{out}} - B_{\text{in}}$$

Stokes theorem yields

$$w_{\text{out}} - w_{\text{in}} = \int_{B_{\text{out}}} \frac{dX}{2\pi R} - \int_{B_{\text{in}}} \frac{dX}{2\pi R} = \int_{\partial \Sigma} \frac{dX}{2\pi R} = \int_{\Sigma} d\left(\frac{dX}{2\pi R}\right) \equiv 0.$$

The Partition Function: Path Integral Derivation

The expression (6.50) may be easily recovered from the path integral over the torus T^2 of modulus τ . We integrate over field configurations given by maps

$$X: T^2 \rightarrow S^1, \tag{6.52}$$

i.e. the functional integration space is $\text{Map}(T^2, U(1))$ which decomposes into homotopy classes labeled by two integers¹² which we identify with the winding numbers $(w, m) \in \mathbb{Z}^2$ around the A - and B - cycles of a chosen marking of T^2 :

$$X(z + 2\pi) = X(z) + 2\pi w R \quad X(z + 2\pi\tau) = X(z) + 2\pi m R. \tag{6.53}$$

Alternatively, consider the one-form on T^2 ,

¹² In general, the homotopy classes of maps $X \rightarrow S^1$ form an Abelian group $[X, S^1]$ isomorphic to $H^1(X, \mathbb{Z})$ cf. **Theorem 4.57** in [15]. This gives $[T^2, S^1] \simeq \mathbb{Z}^2$.

BOX 6.4 - Poisson summation formula

We write $\delta_{\mathbb{Z}}(x)$ for the characteristic distribution of the integer subset $\mathbb{Z} \subset \mathbb{R}$, i.e.

$$\delta_{\mathbb{Z}}(x) \stackrel{\text{def}}{=} \sum_{n=-\infty}^{+\infty} \delta(x - n)$$

with $\delta(x)$ the Dirac δ -function.

Theorem (Poisson summation formula [10]). $\delta_{\mathbb{Z}}(x)$ is its own Fourier transform, that is,

$$\delta_{\mathbb{Z}}(x) = \widehat{\delta_{\mathbb{Z}}}(x) \stackrel{\text{def}}{=} \int_{-\infty}^{+\infty} dp e^{2\pi i p x} \delta_{\mathbb{Z}}(p).$$

Equivalently, let $f: \mathbb{R} \rightarrow \mathbb{C}$ be a function of bounded total variation which is in $L^1(\mathbb{R})$ and whose Fourier transform

$$\widehat{f}(x) = \int_{-\infty}^{+\infty} dp e^{2\pi i p x} f(p)$$

is also in $L^1(\mathbb{R})$: then

$$\sum_{n \in \mathbb{Z}} f(n) = \sum_{m \in \mathbb{Z}} \widehat{f}(m). \quad (\clubsuit)$$

Proof Consider the function $F(x) = \sum_{n \in \mathbb{Z}} f(x + n)$. It is periodic of period 1 and of bounded total variation. Hence it has a Fourier series

$$\begin{aligned} F(x) &= \sum_{m \in \mathbb{Z}} a_m e^{2\pi i m x} \equiv \sum_{m \in \mathbb{Z}} e^{2\pi i m x} \int_0^1 dy e^{-2\pi i m y} F(y) = \\ &= \sum_{m \in \mathbb{Z}} e^{2\pi i m x} \sum_{n \in \mathbb{Z}} \int_0^1 dy e^{-2\pi i m y} f(y + n) = \sum_{m \in \mathbb{Z}} e^{2\pi i m x} \int_{-\infty}^{+\infty} dy e^{-2\pi i m y} f(y) \end{aligned}$$

that is,

$$F(x) = \sum_{m \in \mathbb{Z}} e^{2\pi i m x} \widehat{f}(m).$$

Setting $x = 0$ we get $(\clubsuit)$. Equation (6.49) is Eq. $(\clubsuit)$ for $f(x) = e^{-\pi a x^2 + 2\pi i b x}$ with Fourier transform

$$\widehat{f}(-x) = \int_{-\infty}^{+\infty} dy e^{-\pi a y^2 + 2\pi i b y - 2\pi i x y} = \frac{1}{\sqrt{a}} e^{-\pi(x-b)^2/a}.$$

Note 6.1 The Poisson formula has several fancier versions: the twisted summation, the adelic version, the summation formula over the integers in a totally real number field, etc.; cf. [10].

$$\eta \stackrel{\text{def}}{=} X^* \frac{d\theta}{2\pi} \equiv \frac{dX}{2\pi R} \quad \theta \text{ the usual angular coordinate on } S^1. \quad (6.54)$$

Since $d\theta/2\pi$ is the generator of $H^1(S^1, \mathbb{Z})$, the 1-form η represents an *integral* cohomology class on T^2 , and hence its cohomology class $[\eta]$ may be expanded in a basis of $H^1(T^2, \mathbb{Z}) \simeq \mathbb{Z}^2$ with *integral* coefficients. We represent the basis elements

BOX 6.5 - The $SL(2, \mathbb{Z})$ Hodge norm

We return to the Iwasawa decomposition of $SL(2, \mathbb{R})$ (Eq. (★) in BOX 4.5). As shown there, the upper half-plane $\mathcal{H} = \{\tau \in \mathbb{C} : \text{Im } \tau > 0\}$ is biholomorphic to the coset $SL(2, \mathbb{R})/SO(2)$ with $SO(2)$ acting on the right as in Eq. (★) of that BOX. The modular transformations act on the representative matrices A of a coset in $SL(2, \mathbb{R})/SO(2)$ by multiplication on the left

$$A \mapsto MA, \quad M \in SL(2, \mathbb{Z}) \quad (\diamond)$$

Cartan's totally geodesic embedding [14] is given by

$$\iota_C : SL(2, \mathbb{R})/SO(2) \hookrightarrow SL(2, \mathbb{R}), \quad A \mapsto AA^t \equiv S$$

Note that the image S of a coset $A SO(2) \subset SL(2, \mathbb{R})$ is independent of the chosen representative A , so S is a well-defined symmetric, positive, and unimodular matrix. In fact:

Lemma 6.1 ι_C yields a global isometry between the upper half-plane $\mathcal{H}$ and the space of positive, symmetric, and real 2×2 matrices of unit determinant.

The matrix S yields an alternative way of writing the points in the upper half-space which has the merit of being a *modular tensor*, whereas the parametrization in terms of the coordinate τ realizes modular transformations in a *non-linear* way. Using formulae in the quoted BOX, we have

$$S = \begin{pmatrix} |\tau|^2/\tau_2 & \tau_1/\tau_2 \\ \tau_1/\tau_2 & 1/\tau_2 \end{pmatrix} \quad \text{or} \quad \Omega^t S \Omega = \begin{pmatrix} 1/\tau_2 & -\tau_1/\tau_2 \\ -\tau_1/\tau_2 & |\tau|^2/\tau_2 \end{pmatrix}$$

where $\Omega = -i\sigma_2$ is the 2×2 symplectic matrix. The modular group acts on S as (cf. Eq. (◇))

$$S \mapsto MSM^t, \quad M \in SL(2, \mathbb{Z}).$$

We define the (dual) Hodge norm of a vector in $\mathbb{Z}^2$ as

$$\left\| \begin{pmatrix} m \\ w \end{pmatrix} \right\|_{\text{Hodge}}^2 = (m \ w) \Omega^t S \Omega \begin{pmatrix} m \\ w \end{pmatrix} = \frac{|m - \tau w|^2}{\tau_2}$$

which is manifestly invariant under the $SL(2, \mathbb{Z})$ action

$$S \mapsto MSM^t \quad \text{and} \quad \begin{pmatrix} m \\ w \end{pmatrix} \mapsto M \begin{pmatrix} m \\ w \end{pmatrix}, \quad M \in SL(2, \mathbb{Z}) \quad (\spadesuit)$$

since $M^t \Omega M = \Omega$. The exponent in Eq. (6.50) is just $-\pi R^2/\alpha'$ times the Hodge norm of $(m, w)^t$; the sum over the points in $\mathbb{Z}^2$ is modular-invariant since $SL(2, \mathbb{Z}) \equiv GL^+(2, \mathbb{Z}) \equiv \text{Aut}^+(\mathbb{Z}^2)$.

of $H^1(T^2, \mathbb{Z})$ by the *harmonic* 1-forms on the flat 2-torus T^2 which are Poincaré dual to the A - and B -cycles, so that

$$\left[\frac{dX}{2\pi R} \right] \equiv [\eta] = w \left[\frac{\bar{\tau} dz - \tau d\bar{z}}{2\pi(\bar{\tau} - \tau)} \right] + m \left[\frac{dz - d\bar{z}}{2\pi(\tau - \bar{\tau})} \right], \quad m, w \in \mathbb{Z}. \quad (6.55)$$

Then, after decomposing into form type, the Hodge decomposition [16] of the closed 1-form dX takes the form

$$\partial X(z) = iR \frac{w\bar{\tau} - m}{2\tau_2} dz + \partial \tilde{X}(z), \quad \bar{\partial} X(\bar{z}) = -iR \frac{w\tau - m}{2\tau_2} d\bar{z} + \bar{\partial} \tilde{X}(\bar{z}), \quad (6.56)$$

where $\tilde{X}(z, \bar{z})$ is a global (i.e. univalued) scalar on T^2 yielding the *d-exact part* of the closed differential dX . The action of a map X in the (w, m) topological sector is

$$\begin{aligned} \frac{1}{2\pi\alpha'} \int_{T^2} \partial X \bar{\partial} X &= \frac{8\pi^2\tau_2}{2\pi\alpha'} R^2 \frac{|m - w\tau|^2}{4\tau_2^2} + \frac{1}{2\pi\alpha'} \int_{T^2} \partial \tilde{X} \bar{\partial} \tilde{X} \\ &= \frac{\pi R^2}{\alpha' \tau_2} |m - w\tau|^2 + \frac{1}{2\pi\alpha'} \int_{T^2} \partial \tilde{X} \bar{\partial} \tilde{X} \end{aligned} \quad (6.57)$$

The path integral factorizes into a functional integral over the *d-exact part* $d\tilde{X}$ and a discrete sum over the cohomology class of $[dX]$. The exact part integral is the same path integral as in the non-compact case, producing a factor $2\pi R Z_X(\tau)$, where $2\pi R$ results from the integration over the constant mode of $\tilde{X}$. The sum over the cohomology classes reproduces the sum in Eq. (6.50) which we obtained by Poisson resummation of the Hilbert space trace of the operator formulation.

Vertex Operators. Mutual Locality

By the state/operator correspondence, there should be local operators associated to winding states of the free periodic scalar CFT. In such a state, $\alpha_0 \neq \tilde{\alpha}_0$, so we need to introduce two independent operators x_L and x_R canonically conjugate to the two independent momenta p_L and p_R

$$[x_L, p_L] = [x_R, p_R] = i. \quad (6.58)$$

The field $X(z, \bar{z})$ splits into left- and right-moving parts

$$X(z, \bar{z}) = X_L(z) + X_R(\bar{z}) \quad (6.59)$$

$$X_L(z) = x_L - i \frac{\alpha'}{2} p_L \log z + i \left(\frac{\alpha'}{2} \right)^{1/2} \sum_{m \neq 0} \frac{\alpha_m}{m z^m} \quad (6.60)$$

$$X_R(\bar{z}) = x_R - i \frac{\alpha'}{2} p_R \log \bar{z} + i \left(\frac{\alpha'}{2} \right)^{1/2} \sum_{m \neq 0} \frac{\tilde{\alpha}_m}{m \bar{z}^m}. \quad (6.61)$$

We write k_L and k_R for the eigenvalues of p_L and p_R , respectively. If we restrict to states/operators with $k_L = k_R$, which depend only on the left-right symmetric combination $X_L(z) + X_R(\bar{z})$, we get back the usual non-compact scalar story.

In the general case, we see the left- and right-moving fields $X_L(z)$, $X_R(\bar{z})$ as independent *chiral* scalars. Chiral bosons may appear in QFT only when the spacetime dimension has the form $d = 4k + 2$, the chiral scalar in $d = 2$ being their simplest instance. Chiral bosons are *extremely subtle* QFTs, and (just as the chiral *fermions*) they are often plagued by perturbative and non-perturbative anomalies. We shall encounter other examples later in this book.

We have already informally introduced the 2d chiral scalars in Chap. 2 when we discussed bosonization of 2d Weyl fermions and simply-laced level 1 Kač-Moody (chiral) currents. In particular, we know that the OPEs take the form

$$X_L(z) X_L(w) \sim -\frac{\alpha'}{2} \log(z - w) \quad (6.62)$$

$$X_R(\bar{z}) X_R(\bar{w}) \sim -\frac{\alpha'}{2} \log(\bar{z} - \bar{w}) \quad (6.63)$$

$$X_L(z) X_R(\bar{w}) \sim 0. \quad (6.64)$$

The vertex of the state $|0; k_L, k_R\rangle$ (the eigenstate of p_L , p_R with $N = \tilde{N} = 0$) is¹³

$$V_{k_L, k_R}(z, \bar{z}) =: e^{ik_L X_L(z) + ik_R X_R(\bar{z})}: \quad (6.65)$$

with OPE

$$\begin{aligned} V_{k_L, k_R}(z, \bar{z}) V_{k'_L, k'_R}(w, \bar{w}) &\sim \\ &\sim (z - w)^{\alpha' k_L k'_L / 2} (\bar{z} - \bar{w})^{\alpha' k_R k'_R / 2} V_{(k_L + k'_L), (k_R + k'_R)}(w, \bar{w}). \end{aligned} \quad (6.66)$$

Since the field $X(z, \bar{z})$ is not single valued on Σ , a priori there may be branch cuts in various OPEs. However physical consistency requires the vertices in the *actual* CFT operator algebra $\mathfrak{A}$ to have single-valued OPEs, i.e. its operators must be all *mutually local*. The quantization of left- and right momenta, Eq. (6.46), enforces this condition: the net phase we pick up in (6.66) when z goes around w is

$$\exp[\pi i \alpha' (k_L k'_L - k_R k'_R)] \equiv \exp[2\pi i (nw' + wn')] = 1, \quad (6.67)$$

where we used Eq. (6.46) which yield

$$\alpha' (k_L k'_L - k_R k'_R) \equiv 2(nw' + n'w). \quad (6.68)$$

Cocycle Factors The OPE (6.66) has a problem: if we interchange $z \leftrightarrow w$ and $k \leftrightarrow k'$ the LHS is symmetric while the RHS picks a phase $\exp[\pi i (nw' + wn')]$, so it changes sign when $nw' + wn'$ is *odd*. We already know how to fix this problem (already encountered in Sect. 2.9); we need to introduce *cocycle factors* which depend only on the zero-mode operators, i.e. on the center of mass left- and right-momentum

¹³ As in Sect. 2.9 the vertex should also be supplemented by a “cocycle” factor, see below.

operators p_L and p_R . The precise formula is

$$V_{k_L, k_R}(z, \bar{z}) = \exp[\pi i(k_L - k_R)(p_L + p_R)\alpha'/4] : e^{ik_L X_L(z) + ik_R X_R(\bar{z})} : \quad (6.69)$$

Correlators The non-compact scalar correlators (4.24) factorize into holomorphic and anti-holomorphic parts. By the Wick theorem, we may simply replace

$$\prod_{i < j} |z_{ij}|^{\alpha' k_i k_j} \longrightarrow \prod_{i < j} z_{ij}^{\alpha' k_{Li} k_{Lj}/2} \bar{z}_{ij}^{-\alpha' k_{Ri} k_{Rj}/2}, \quad (6.70)$$

and also replace the momentum Dirac-delta $2\pi\delta(\sum_i k_i)$ arising from the non-compact integral over x_0 with Kronecker deltas arising from Poisson resummation:

$$2\pi R \delta_{\sum_i n_i, 0} \delta_{\sum_i w_i, 0} \quad (6.71)$$

which express conservation of both compact momentum and total winding number.

★ **Marginal Deformations** An operator $O(z, \bar{z}) \in \mathfrak{A}$ is *marginal* iff its addition to the action $S \rightsquigarrow S + \delta g \int d^2 z O(z, \bar{z})$ deforms the QFT while keeping it conformal invariant. The marginal couplings g are local coordinates in the (connected) conformal “manifold”¹⁴ $\mathcal{S}$, i.e. the space which parametrize the continuous family of CFTs. The central charge c is *constant* in $\mathcal{S}$ [18]. A *Hermitian* operator is *marginal* iff [17]: (i) its weights are $(h, \bar{h}) = (1, 1)$ and (ii) the OPE $O(z, \bar{z})O(w, \bar{w})$ does not contain any primary with $(h, \bar{h}) = (1, 1)$. For a compact scalar at generic radius R , these conditions hold for $\partial X \bar{\partial} X$; the addition of this operator to the Lagrangian can be undone by a rescaling of the field X , i.e. by a variation of R . The theory is conformal for all R , so $\partial X \bar{\partial} X$ is obviously marginal. However, for *special* R , there are additional marginal operators. There are three special values: $R = \sqrt{\alpha'}$, $R = \sqrt{\alpha'}/2$, and $R = 2\sqrt{\alpha'}$. The first value is related to the phenomenon of *gauge symmetry enhancement*, and will be discussed in detail in Sect. 6.4. The other two special radii yield equivalent realizations of the CFT which describes the Kosterlitz–Thouless multi-critical transition point of the XY model [19]. We shall comment on it in Sect. 6.6.2.

6.3 ★ Bosonization: Riemann Identities for Partition Functions

We return to the partition function for a complex left-moving chiral fermion λ subject to the general periodic b.c. (5.65) where $a \equiv \alpha/2$ and $b \equiv \beta/2$ are arbitrary real numbers

$$\text{Tr}_{2a} [q^{L_0 - c/24} e^{2\pi i b J}] = \frac{\vartheta \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] (0, \tau)}{\eta(\tau)}. \quad (6.72)$$

For future convenience, we define the allied functions in two variables

¹⁴ Although the space $\mathcal{S}$ is smooth at a generic point, there are special points, called *multi-critical* where $\mathcal{S}$ has singularities [17].

$$Z_{\beta}^{\alpha}(y, \tau) \stackrel{\text{def}}{=} e^{\pi i y^2 / \tau} \frac{\vartheta \left[\begin{smallmatrix} \alpha/2 \\ \beta/2 \end{smallmatrix} \right] (y, \tau)}{\eta(\tau)}, \quad \alpha, \beta \in \{0, 1\}, \quad q \equiv e^{2\pi i \tau}, \quad \tau \in \mathcal{H}, \quad y \in \mathbb{C} \quad (6.73)$$

which at $y = 0$ give back the Fermi torus partition function in the various spin-structures $Z_{\beta}^{\alpha}(\tau)$. The physical interpretation of the new functions is $(a, b \in \mathbb{R})$

$$Z_{\beta}^{\alpha}(b + a\tau, \tau) = \text{Tr}_{2a+\alpha} [q^{L_0 - c/24} e^{2\pi i (b + \beta/2) J_a}], \quad (6.74)$$

where $J_a \equiv J - a$ is the $U(1)$ charge subtracted so that the a -twisted vacuum (i.e. Fermi sea) $|a\rangle$ has J_a -charge zero.

We prove a set of identities for the functions $Z_{\beta}^{\alpha}(y, \tau)$ called the *Riemann relations* which will be used several times in this book. From a CFT standpoint, these identities just reflect bosonization and subsequent re-fermionization of the level 1 $\text{Spin}(8)$ current algebra already discussed in Sect. 2.9.3. We start by giving a CFT interpretation of Jacobi triple-product identity in BOX 5.5.

Jacobi Triple Product versus Bosonization

The (elliptic) ϑ -functions with characteristic (cf. BOX 5.5) can be written either as a q -series or as an infinite product (Jacobi triple product). In Eq. (5.71), the infinite product was interpreted as the partition function of the Fermi free gas of oscillator modes of a Weyl fermion $\lambda(z)$: each factor in the infinite product is the Fermi partition function of a mode of $\lambda(z)$. We wish to give a physical interpretation of the series expression:

$$Z_0^0(\tau) = \frac{\vartheta \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} \right] (0, \tau)}{\eta(\tau)} = \frac{q^{-1/24}}{\prod_{n \geq 1} (1 - q^n)} \sum_{n \in \mathbb{Z}} q^{\frac{1}{2}n^2} \quad (6.75)$$

Comparing with the holomorphic factor in (6.48) we see that $Z_0^0(\tau)$ is the partition function of a left-moving chiral scalar with $p_L = n$, $p_R = 0$, i.e. with momenta taking value in the lattice $\mathbb{Z}$ which—according to Sect. 2.9.2—is precisely the bosonization of a free complex fermion $\lambda(z)$. Thus the Jacobi identity, which equates the sum and the product, is just the statement that the partition function of the complex free fermion can be equally well computed as the partition function of the chiral scalar which bosonizes it. In other words, the identity is a combinatorial proof of bosonization.

Re-Fermionization and Riemann Relations We recall two formulae from BOX 5.5:

$$\vartheta \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] (z|\tau) \stackrel{\text{def}}{=} \sum_{n \in \mathbb{Z}} \exp[\pi i (n + a)^2 \tau + 2\pi i (n + a)(z + b)] \quad (6.76)$$

$$\vartheta \left[\begin{smallmatrix} a + m \\ b + n \end{smallmatrix} \right] (z|\tau) = e^{2\pi i n a} \vartheta \left[\begin{smallmatrix} a \\ b \end{smallmatrix} \right] (z|\tau) \quad a, b \in \mathbb{R}. \quad (6.77)$$

When the characteristic is half-integral, we use a special notation:

$$\theta_{\alpha, \beta}(z, \tau) \equiv \vartheta \left[\begin{smallmatrix} \alpha/2 \\ \beta/2 \end{smallmatrix} \right] (z|\tau) \quad \alpha, \beta \in \{0, 1\}. \quad (6.78)$$

There is a large family of quartic identities for ϑ -functions. From the physical viewpoint, they arise from a special property of the 2d $SO(8)$ current algebra. Its zero-mode Lie subalgebra $\mathfrak{spin}(8)$ has a $\mathfrak{S}_3$ group of outer automorphisms, called *triality*, induced by the $\mathfrak{S}_3$ automorphisms of the D_4 Dynkin graph, see Eq. (2.528); the triality automorphism permutes the vector representation and the two spinor representations. Triality extends to an automorphisms of the full $\text{Spin}(8)$ current algebra which leaves invariant the Sugawara energy-momentum tensor.

Physically the Riemann ϑ -relations express the invariance of the torus path integral under *re-fermionization*; we bosonize the $SO(8)$ fermions ψ^i , then construct the spin operators S_{α} , check that they satisfy the free fermion OPE and re-fermionize them cf. Sect. 2.9.3. Triality is equivalent to the following chain of isomorphisms of $\text{Spin}(8)$ GSO-projected lattices (cf. Sects. 2.9.2 and 2.9.3).

$$I_4 \equiv (o) + (v) \simeq (o) + (s) \simeq (0) + (c). \quad (6.79)$$

Recall from **BOX 2.10** that a *lattice* is a free Abelian group of finite rank, $\Gamma = \oplus_i \mathbb{Z} e_i$, with a quadratic form $A(n_i e_i) \equiv A_{ij} n_i n_j$ where $A_{ij} = A_{ji} \in \mathbb{Z}$. Two rank- r lattices are *isomorphic*, $\Gamma_1 \simeq \Gamma_2$, iff their quadratic forms are $\mathbb{Z}$ -equivalent, i.e. there is a matrix $S \in GL(r, \mathbb{Z})$ which is an isometry of the quadratic forms A_1, A_2 , that is,

$$A_2 = S^t A_1 S. \quad (6.80)$$

In (6.79), I_4 is the “standard” rank-4 lattice $\mathbb{Z}^4$ with the quadratic form $\sum_{i=1}^4 n_i^2$. Two positive-definite lattices Γ_1, Γ_2 are isomorphic if and only if they have the same ϑ -constants

$$\sum_{n_i \in \Gamma_1} q^{\frac{1}{2} A_{1,ij} n_i n_j} = \sum_{n_i \in \Gamma_2} q^{\frac{1}{2} A_{2,ij} n_i n_j / 2} \quad |q| < 1. \quad (6.81)$$

More generally, we can consider ϑ -functions with characteristics

$$\begin{aligned} & \sum_{n_i \in \Gamma_2} \exp[\pi i \tau A_{2,ij} (n_i + a_i)(n_j + a_j) + 2\pi i (n_i + a_i)(z_i + b_i)] = \\ & = \sum_{n_i \in \Gamma_1} \exp\left[\pi i \tau A_{1,ij} (n_i + (Sa)_i)(n_j + (Sa)_j) + 2\pi i (n_i + (Sa)_i)((S^{-t}z)_i + (S^{-t}b)_i)\right] \end{aligned} \quad (6.82)$$

Let us apply this equality to the re-fermionization isomorphism $(o) + (s) \simeq I_4$. The RHS becomes

$$\begin{aligned} & \prod_{i=1}^4 \sum_{n \in \mathbb{Z}} \exp[\pi i \tau (n_i + (Sa)_i)^2 + 2\pi i (n_i + (Sa)_i)((S^{-t}z)_i + S^{-t}b_i)] = \\ & = \prod_{i=1}^4 \vartheta \left[\begin{matrix} (Sa)_i \\ (S^{-t}b)_j \end{matrix} \right] ((S^{-t}z)_i, \tau) \end{aligned} \quad (6.83)$$

i.e. the product of 4 ϑ ’s with arguments $(S^{-t}z)_i$ and specific characteristics. We consider the LHS; as in the bosonization procedure, we see $(o) + (s)$ as the sublattice of $(\frac{1}{2}\mathbb{Z})^4$ of vectors of the form

$$(o) \quad (n_1, n_2, n_3, n_4) \quad n_i \in \mathbb{Z}, \quad \sum n_i \in 2\mathbb{Z} \quad (6.84)$$

$$(s) \quad (n_1 + \frac{1}{2}, n_2 + \frac{1}{2}, n_3 + \frac{1}{2}, n_4 + \frac{1}{2}) \quad n_i \in \mathbb{Z}, \quad \sum n_i \in 2\mathbb{Z} \quad (6.85)$$

rather than as a copy of $\mathbb{Z}^4$. With this normalization $S \in GL(4, \frac{1}{2}\mathbb{Z})$

$$S = \frac{1}{2} \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & -1 & +1 \end{pmatrix} \quad (6.86)$$

and the isomorphism $(o) + (s) \simeq I_4$ is just the statement that S is orthogonal

$$S^t S = 1, \quad S^{-t} = S. \quad (6.87)$$

Hence the LHS of Eq. (6.82) is

$$\begin{aligned} & \sum_{m_i \in (o)+(v)} e^{\pi i \tau (m_i + a_i)^2 + 2\pi i (m_i + a_i)(z_i + b_i)} = \\ &= \sum_{\substack{n_i \in \mathbb{Z} \\ \sum_i n_i \in 2\mathbb{Z}}} e^{\pi i \tau (n_i + a_i)^2 + 2\pi i (n_i + a_i)(z_i + b_i)} + \sum_{\substack{n_i \in \mathbb{Z} \\ \sum_i n_i \in 2\mathbb{Z}}} e^{\pi i \tau (n_i + 1/2 + a_i)^2 + 2\pi i (n_i + 1/2 + a_i)(z_i + b_i)} \end{aligned} \quad (6.88)$$

to implement the restriction of the sum to $\sum_i n_i \in 2\mathbb{Z}$ we insert in the sum the GSO projector $\frac{1}{2} (1 + e^{\pi i \sum_i n_i})$. Then twice the LHS of Eq. (6.82) is

$$\begin{aligned} & \sum_{\alpha=0,1} \sum_{n_i \in \mathbb{Z}} e^{\pi i \tau (n_i + \alpha/2 + a_i)^2 + 2\pi i (n_i + \alpha/2 + a_i)(z_i + b_i)} + \\ & + \sum_{\alpha=0,1} \sum_{n_i \in \mathbb{Z}} e^{\pi i \tau (n_i + \alpha/2 + a_i)^2 + 2\pi i (n_i + \alpha/2 + a_i)(z_i + 1/2 + b_i) - \pi i \sum_i a_i} = \\ &= \sum_{\alpha=0,1} \prod_{i=1}^4 \vartheta \left[\begin{smallmatrix} a_i + \frac{\alpha}{2} \\ b_i \end{smallmatrix} \right] (z_i, \tau) + e^{-\pi i \sum_i a_i} \sum_{\alpha=0,1} \prod_{i=1}^4 \vartheta \left[\begin{smallmatrix} a_i + \frac{\alpha}{2} \\ b_i + \frac{1}{2} \end{smallmatrix} \right] (z_i, \tau). \end{aligned} \quad (6.89)$$

Equating *one-half* of the RHS of this equation to the RHS of (6.83), we get a huge family of quartic theta identities called the *Riemann ϑ -relations* [20]. For the purpose of bosonization, the most relevant identities are the ones with $(Sa)_i$ (resp. $(S^{-t}b)_i$) equal to $\alpha/2$ (resp. $\beta/2$) for all i . This gives

$$a_i = \alpha \delta_{i,1}, \quad b_i = \beta \delta_{i,1}, \quad \alpha, \beta \in \{0, 1\}. \quad (6.90)$$

This particular specialization of the RHS of Eq. (6.89) reads

$$\prod_{i=1}^4 \vartheta \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix} \right] (z_i, \tau) + e^{\pi i \alpha} \prod_{i=1}^4 \vartheta \left[\begin{smallmatrix} 0 \\ \frac{1}{2} \end{smallmatrix} \right] (z_i, \tau) + e^{\pi i \beta} \prod_{i=1}^4 \vartheta \left[\begin{smallmatrix} \frac{1}{2} \\ 0 \end{smallmatrix} \right] (z_i, \tau) + e^{\pi i (\alpha + \beta)} \prod_{i=1}^4 \vartheta \left[\begin{smallmatrix} \frac{1}{2} \\ \frac{1}{2} \end{smallmatrix} \right] (z_i, \tau) \quad (6.91)$$

where we used Eq. (6.77). Using the notation (6.78), we get the re-bosonization identity in the form

$$\begin{aligned} \prod_{i=1}^4 \theta_{\alpha, \beta}((Sz)_i, \tau) &= \frac{1}{2} \prod_{i=1}^4 \theta_{0,0}(z_i, \tau) + \frac{1}{2} (-1)^\alpha \prod_{i=1}^4 \theta_{0,1}(z_i, \tau) + \\ &+ \frac{1}{2} (-1)^\beta \prod_{i=1}^4 \theta_{1,0}(z_i, \tau) + \frac{1}{2} (-1)^{\alpha + \beta} \prod_{i=1}^4 \theta_{1,1}(z_i, \tau) \end{aligned} \quad (6.92)$$

Since $S^t S = 1$, $\exp(i\pi \sum_i y_i^2 / \tau) \equiv \exp(i\pi \sum_i (Sy)_i^2 / \tau)$. Then we may rewrite (6.92) in terms of the functions $Z_\beta^\alpha(y, \tau)$ defined in Eq. (6.73)

$$\begin{aligned} \prod_{i=1}^4 Z_\beta^\alpha((Sy)_i, \tau) &= \frac{1}{2} \left[\prod_{i=1}^4 Z_0^0(y_i, \tau) + (-1)^\alpha \prod_{i=1}^4 Z_1^0(y_i, \tau) + \right. \\ &\quad \left. + (-1)^\beta \prod_{i=1}^4 Z_0^1(y_i, \tau) + (-1)^{\alpha + \beta} \prod_{i=1}^4 Z_1^1(y_i, \tau) \right] \end{aligned} \quad (6.93)$$

The equation may be written in terms of traces on the Hilbert space of 8 MW fermions as

$$\begin{aligned}
\text{Tr}_\alpha \left[(-1)^{\beta F} e^{2\pi i (Sz)_i J_i} q^{L_0 - c/24} \right] &= \\
&= \sum_{\alpha'=0}^1 (-1)^{\beta \alpha'} \text{Tr}_{\alpha'} \left[\frac{1 + (-1)^{F+\alpha}}{2} e^{2\pi i z_i J_i} q^{L_0 - c/24} \right]
\end{aligned} \tag{6.94}$$

where the trace in the RHS (resp. lhs) is over the Hilbert space of the *original* fermions (resp. of the *re-fermionized* ones). More generally, we may consider arbitrary half-integer characteristics $(a_i, b_i) \in (\frac{1}{2}\mathbb{Z})^8$ subjected only to the conditions $\sum_i a_i, \sum_i b_i \in \mathbb{Z}$ which guarantee that the dual characteristics $((Sa)_i, (S)_i) \in (\frac{1}{2}\mathbb{Z})^8$ are also half-integral.

Exercise 6.4 Write the identity for $\prod_{i=1}^4 \theta_{\alpha_i, \beta_i}((Sz)_i, \tau)$ with $\sum_i \alpha_i, \sum_i \beta_i \in 2\mathbb{Z}$.

6.4 T -Duality in Closed Strings

For a general reference about target-space dualities, see the nice review [21].

We consider the critical ($d = 26$) closed oriented bosonic string on

$$\text{target space-time} \equiv \mathbb{R}^{24,1} \times S^1, \quad ds^2 = \eta_{MN} dX^M dX^N, \tag{6.95}$$

where the last coordinate X^{25} is periodic of period $2\pi R$ while X^μ is an ordinary non-compact coordinate for $\mu = 0, \dots, 24$. The IR physics is effectively 25-dimensional.

Mass Formulae

The mass-shell conditions from the 25-dimensional perspective are

$$m^2 = -k_\mu k^\mu = \begin{cases} (k_L^{25})^2 + \frac{4}{\alpha'}(N-1) \\ (k_R^{25})^2 + \frac{4}{\alpha'}(\tilde{N}-1) \end{cases} \tag{6.96}$$

that is, using Eq. (6.46),

$$m^2 = \frac{n^2}{R^2} + \frac{w^2 R^2}{\alpha'^2} + \frac{2}{\alpha'}(N + \tilde{N} - 2) \tag{6.97}$$

$$0 = nw + N - \tilde{N}. \tag{6.98}$$

There are four terms in the mass-squared: the compact momentum squared, the tension energy of the winding string, the oscillators, and the zero-point energy, i.e. the weight $h + \tilde{h} \equiv -2$ of the ghost sea $|c\tilde{c}\rangle$. As always, we can count states keeping only transverse oscillators by the no-ghost theorem. Note that, in the winding sectors $w \neq 0$, the left-right matching condition has the modified form (6.98).

Generic Radius: Gauge Fields

Let us first determine the massless spectrum *at a generic value* of the compactification radius R . At generic R , the only way a state may be massless is if

$$n = w = 0 \quad N = \tilde{N} = 1. \tag{6.99}$$

These are the same 24^2 massless states as in the non-compact theory, but now we organize them in representations of the effective Lorentz group $SO(24, 1)$

$$\begin{aligned} & \alpha_{-1}^{\mu} \tilde{\alpha}_{-1}^{\nu} |0; k\rangle & (\alpha_{-1}^{\mu} \tilde{\alpha}_{-1}^{25} + \alpha_{-1}^{25} \tilde{\alpha}_{-1}^{\mu}) |0; k\rangle \\ & (\alpha_{-1}^{\mu} \tilde{\alpha}_{-1}^{25} - \alpha_{-1}^{25} \tilde{\alpha}_{-1}^{\mu}) |0; k\rangle & \alpha_{-1}^{25} \tilde{\alpha}_{-1}^{25} |0; k\rangle \end{aligned} \quad (6.100)$$

The first state represents a 25-dimensional graviton plus dilaton and 2-form field $B_{\mu\nu}$; the second one a graviton with one internal index, $\mathbf{G}_{\mu 25}$, that is the Kaluza–Klein vector; the third state is the vector from the 2-form field $B_{\mu} \equiv \mathbf{B}_{\mu 25}$. The last state is a scalar: it is the modulus σ associated to fluctuations of the radius R of the S^1 factor space in (6.3). This is the same massless spectrum we found studying the low-energy effective field theory.

The massive states are charged under the $U(1) \times U(1)$ gauge symmetry. The Kaluza–Klein $U(1)$ is the group of translation in the S^1 fiber, $X^{25} \rightarrow X^{25} + \text{const.}$, and its quantized charge is the compact momentum p_{25} . The conserved charge associated to $U(1)$ symmetry gauged by the vector B_{μ} is the winding number w : for a detailed derivation see **BOX 6.6**.

The *winding gauge charge* is a stringy feature: in a *typical* field theory there is *no* state carrying the winding-number charge even when a massless 2-form field $B_{\mu\nu}$ is present. However in the (highly non-generic) effective field theories, which do arise as low-energy limits of a quantum gravity and contain a massless 2-form gauge field B , there are stable *solitonic* extended objects carrying the associated winding charge. Indeed, in a consistent quantum gravity, there are objects carrying all possible values of the conserved gauge charges; this is one *swampland conjecture* [22].

Enhanced Gauge Symmetry

In the previous paragraph, we omitted the states which are massless *only* for *special* values of R . The richest massless spectrum is obtained when $R = \sqrt{\alpha'}$; in this case

$$k_{L,R}^{25} = \frac{1}{\sqrt{\alpha'}} (n \pm w), \quad (6.101)$$

and the condition for a massless state becomes

$$(n + w)^2 + 4N = (n - w)^2 + 4\tilde{N} = 4. \quad (6.102)$$

In addition to the generic solution $n = w = 0$, $N = \tilde{N} = 1$ we have also

$$n = w = \pm 1, \quad N = 0, \quad \tilde{N} = 1, \quad \text{and} \quad n = -w = \pm 1, \quad N = 1, \quad \tilde{N} = 0, \quad (6.103)$$

$$n = \pm 2, \quad w = N = \tilde{N} = 0, \quad \text{and} \quad w = \pm 2, \quad n = N = \tilde{N} = 0. \quad (6.104)$$

The states (6.103) include four new massless gauge bosons with vertex operators¹⁵

¹⁵ As always, the definition of the operation involves a cocycle which we omit for simplicity. **Warning:** here and below we omit the ghost factor $c\tilde{c}$ in the vertices to simplify the formulae.

$$:\bar{\partial} X^\mu e^{ikX} \exp[\pm 2i X_L^{25}/\sqrt{\alpha'}]: \quad : \partial X^\mu e^{ikX} \exp[\pm 2i X_R^{25}/\sqrt{\alpha'}]: \quad (6.105)$$

BOX 6.6 - Winding number as the 2-form gauge charge

The coupling of the string world-sheet Σ with the 2-form field

$$\frac{1}{2\pi\alpha'} \int_{\Sigma} B \equiv \frac{1}{4\pi\alpha'} \int_{\Sigma} d^2\sigma \epsilon^{\alpha\beta} B_{MN} \partial_{\alpha} X^M \partial_{\beta} X^N \quad (\clubsuit)$$

is the generalization of 1-form field coupling to the world-line L of a charged particle

$$e \int_L A \equiv e \int_L A_{\mu} dx^{\mu} \quad (\spadesuit)$$

with the obvious modification for dimension 2. We can write $(\spadesuit)$ as an integral over spacetime

$$\int_M d^n x j^{\mu}(x) A_{\mu}(x)$$

where the *electric current* $j^{\mu}(x)$ is $j^{\mu}(x) = e \int_L dt \partial_t x^{\mu} \delta(x - X(t))$ with $X(t)$ the world-line of the particle. The electric charge of the particle is

$$Q_e \stackrel{\text{def}}{=} \int d^{n-1} \mathbf{x} j^0(\mathbf{x}; 0)$$

Likewise, we may write the 2-form/string coupling $(\clubsuit)$ as an integral over spacetime

$$\frac{1}{2} \int_M d^n x j^{MN}(x) B_{MN}(x)$$

where the $(n-2)$ -form $*j$ is the current with support on the image of the worldsheet Σ , $X(\Sigma) \subset M$ which is Poincaré dual to the 2-cycle $X(\Sigma)$ [23], i.e.

$$j^{MN}(x) = \frac{1}{2\pi\alpha'} \int_{\Sigma} d^2\sigma \epsilon^{\alpha\beta} \partial_{\alpha} X^M \partial_{\beta} X^N \delta^n(x - X(\sigma)).$$

Integrating this current at fixed time gives the corresponding charge

$$\begin{aligned} Q^M &= \int d^{n-1} \mathbf{x} J^{M0}(\mathbf{x}) = \frac{1}{2\pi\alpha'} \int_{\Sigma} d^2\sigma \epsilon^{\alpha\beta} \partial_{\alpha} X^M \partial_{\beta} X^N \delta(X^0(\sigma) - t) = \\ &= \frac{1}{2\pi\alpha'} \int_{\Sigma} dX^M \wedge dX^0 \delta(X^0 - t) = \frac{1}{2\pi\alpha'} \int_{\Sigma} dX^M \wedge d\Theta(X^0 - t) = \frac{1}{2\pi\alpha'} \int_{\ell_t} dX^M, \end{aligned}$$

where the curve ℓ_t is the position of the string in space at physical time t

These states have internal momentum and winding number, so they carry KK and B -field gauge charges. The only consistent IR theory of charged massless vectors is Yang-Mills theory, so the new gauge bosons should combine with the old ones to form a non-Abelian gauge system. We use the basis

$$\partial X^{25} \bar{\partial} X^\mu, \quad \partial X^\mu \bar{\partial} X^{25} \quad (6.106)$$

for the vertices of the two generic massless vectors. The first generic vector couples to the charge k_L^{25} under which the 2 states in the left side of (6.105) carry charges ± 1 and the 2 states on the right are neutral. The second generic vector couples to k_R^{25} , and hence to the 2 states on the right side of (6.105). This identifies the gauge group to be $SU(2) \times SU(2)$ with the three vectors involving $\bar{\partial} X^\mu$ in their vertices gauging the first $SU(2)$ group and the three vectors involving ∂X^μ gauging the second $SU(2)$.

To exhibit, say, the first $SU(2)$ gauge symmetry, consider the three $(1, 0)$ (chiral, left-moving) currents

$$j^1(z) = : \cos[2\alpha'^{-1/2} X_L^{25}(z)] : \quad (6.107)$$

$$j^2(z) = : \sin[2\alpha'^{-1/2} X_L^{25}(z)] : \quad (6.108)$$

$$j^3(z) = i \partial X_L^{25} / \alpha'^{1/2} \quad (6.109)$$

which are normalized so that their OPE read

$$j^i(z) j^j(0) \sim \frac{\delta^{ij}}{2z^2} + i \frac{\epsilon^{ijk}}{z} j^k(0), \quad (6.110)$$

which is the $SU(2)$ Kač-Moody current algebra at level 1

$$j^i(z) = \sum_{m \in \mathbb{Z}} \frac{j_m^i}{z^{m+1}}, \quad [j_m^i, j_n^j] = \frac{m}{2} \delta^{ij} \delta_{m+n,0} + i \epsilon^{ijk} j_{m+n}^k. \quad (6.111)$$

In other words, the (say) left-moving part X_L^{25} of a compact scalar at the radius $R = \sqrt{\alpha'}$ is the chiral scalar which bosonizes *à la* Frenkel-Kač-Segal the level 1 $SU(2)$ current algebra; cf. Chap. 2.

Of course there is a similar algebra of anti-holomorphic $(0, 1)$ currents making a second right-moving copy of the $SU(2)$ current algebra. The existence of this $SU(2) \times SU(2)$ gauge symmetry for a special radius R is a first hint that string theory sees the spacetime geometry in a way very different from field theory where only the Abelian $U(1) \times U(1)$ gauge symmetry is realized for all R .

Higgs Mechanism

We move a little bit away from the $SU(2) \times SU(2)$ radius $R = \sqrt{\alpha'}$. The gauge bosons with vertices (6.105) now acquire a mass

$$m = \frac{|R^2 - \alpha'|}{R \alpha'} \approx \frac{2}{\alpha'} |R - \alpha'^{1/2}| \quad \text{for } R \approx \sqrt{\alpha'} \quad (6.112)$$

For R close to $\sqrt{\alpha'}$, the mass of the vectors (6.105) is much smaller than the string scale $1/\sqrt{\alpha'}$, and we may understand what is going on in terms of the low-energy field theory. In field theory, there is only one mechanism to give a mass to a gauge

boson: *spontaneously symmetry breaking*. At $R = \alpha'^{1/2}$ there are 10 massless scalars: the dilaton, the modulus σ , the four scalars in Eq. (6.103) where a 25-oscillator is excited, and the four states (6.104). The last nine scalars have vertices of the form

$$: j^i(z) \tilde{j}^j(\bar{z}) e^{ikX(z, \bar{z})} : \quad (6.113)$$

where the index i (resp. j) is a vector index for the left-moving (resp. right-moving) $SU(2)$ current algebra. These 9 scalars then transform in the $(\mathbf{3}, \mathbf{3})$ of $SU(2) \times SU(2)$. The modulus $\sigma(X^\mu)$ associated to fluctuations of R enters in the world-sheet coupling

$$\frac{1}{\pi\alpha'} \int_{\Sigma} \exp[2\sigma(X^\mu)] \partial X^{25} \bar{\partial} X^{25}, \quad (6.114)$$

and corresponds to the vertex $: j^3 \tilde{j}^3 \exp(ikX) :$ the $i = j = 3$ component of (6.113). Moving away from the $SU(2) \times SU(2)$ radius means giving a v.e.v. to this scalar, and hence to break the gauge symmetry to the subgroup $U(1) \times U(1)$ which leaves invariant its v.e.v. $\langle \sigma \rangle \approx (R - \alpha'^{1/2})/\alpha'^{1/2}$. Near the $SU(2) \times SU(2)$ radius the mass is linear in the breaking parameter $|\langle \sigma \rangle|$ as it should.

Note 6.2 The self-dual radius is a multi-critical point with several marginal operators, in fact a $S^2 \times S^2$ -family of them: $c^i \tilde{c}^j \tilde{j}_i(z) j_j(\bar{z})$ where the 3-vectors $c^i, \tilde{c}^j$ belong to the unit sphere. The symmetry $SU(2)_L \times SU(2)_R$ acts transitively on this family, so all marginal deformations give *equivalent CFTs*, namely the compact scalar at nearby radius R .

6.4.1 *T*-Duality for a Compact Scalar

From the mass formula

$$m^2 = \frac{n^2}{R^2} + \frac{w^2 R^2}{\alpha'^2} + \frac{2}{\alpha'}(N + \tilde{N} - 2) \quad (6.115)$$

we see that as $R \rightarrow \infty$ (the *decompactification limit*) the winding states become infinitely massive, while the momentum spectrum becomes continuous. In the opposite limit, $R \rightarrow 0$ the states with compact momentum get infinitely massive, whereas the winding modes have a continuous spectrum. Thus as the radius goes to zero, we also get a mass spectrum typical of a non-compact dimension. This is quite different from what happens in field theory where there are modes carrying compact momentum but not states¹⁶ which become light as $R \rightarrow 0$.

In fact the $R \rightarrow 0$ and $R \rightarrow \infty$ limits are *physically equivalent*. The spectrum (6.115) is invariant under

¹⁶ Except for string-like solitons. In this case, the field theory behaves “effectively” as a string theory but only for R larger than a certain critical scale of the QFT.

$$R \leftrightarrow \alpha'/R \quad n \leftrightarrow w. \quad (6.116)$$

The equivalence extends to the interactions as well; in Sect. 6.4.2, we present a proof of this fact which is valid for all amplitudes and in a more general context. From Eq. (6.46), we see that (6.116) is equivalent to

$$p_L^{25} \leftrightarrow p_L^{25}, \quad p_R^{25} \leftrightarrow -p_R^{25}. \quad (6.117)$$

Consider the theory at radius R and let

$$X^{25}(z, \bar{z}) \equiv X_L^{25}(z) + X_R^{25}(\bar{z}) \quad (6.118)$$

be the decomposition of the compact scalar field X^{25} in its left- and right-moving parts. Extending (6.117) to non-zero-modes, we define the *dual scalar field* as

$$X'^{25}(z, \bar{z}) = X_L^{25}(z) - X_R^{25}(\bar{z}). \quad (6.119)$$

The field $X'^{25}(z, \bar{z})$ has the same OPE and energy-momentum tensor as $X^{25}(z, \bar{z})$. The only modification in the CFT when we replace $X^{25}(z, \bar{z})$ with $X'^{25}(z, \bar{z})$ is Eq. (6.117) which maps the spectrum at R to the one at the *dual radius*

$$R' = \alpha'/R. \quad (6.120)$$

The world-sheet theories at two dual radii R and α'/R are then the *same* CFT written in terms of the fields $X^{25}(z, \bar{z})$ and, respectively, $X'^{25}(z, \bar{z})$. We stress that the field redefinition from the original to the dual scalar field is highly *non-local*.

The physical on-shell backgrounds for the bosonic string are the 2d CFTs with $c = 26$, and different ways of representing a given CFT yield the *same* physical background. Hence the two compactifications at radii R and α'/R are not just two different situations with the same physics, but they are literally the *same* physical configuration, and the possibility of writing them in two different ways corresponds to a redundancy of the formalism, i.e. they are identified up to some suitable “gauge equivalence”. In the next paragraph, we shall revisit this issue from a more conventional viewpoint.

The quantum equivalence of the compact scalar at radii R and $R' \equiv \alpha'/R$ is known as *T -duality*.

The physical equivalence of the $R \rightarrow 0$ and $R \rightarrow \infty$ limits is in sharp contrast with what happens in the point-particle case, and is yet another hint that string theory sees short-distance geometry in a very different way from field theory. The space of inequivalent backgrounds is the half-line $R \geq \sqrt{\alpha'}$; there is no physical radius smaller than the *self-dual radius*

$$R_{\text{self-dual}} = R_{SU(2) \times SU(2)} = \alpha'^{1/2}. \quad (6.121)$$

Equation (6.119) may be phrased in a way which is often convenient; T -duality is a spacetime parity operation which affects only the right-movers, leaving the left-movers invariant. Several remarks are in order:

Remark 6.1 In Chap. 5, we noticed that in the closed string, there is no UV region, in the sense that we may always reinterpret it as an IR limit by a modular transformation. Here we see another manifestation of the same stringy physical principle; the small length limit $R \rightarrow 0$ may be reinterpreted as a large length limit $R' \rightarrow \infty$. These two observations are directly related. From Eq. (♠) in BOX 6.5, we see that the modular transformation $S \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ has the effect

$$n \rightarrow n' \equiv -w, \quad w \rightarrow w' \equiv n, \quad (6.122)$$

i.e. S acts on the string quantum numbers as T -duality.

Remark 6.2 In QFT, the RG flow from the UV to the IR is *irreversible* [18, 24], and going to the IR we “lose” information. This entails that the UV completion of a given IR effective theory, if it exists, is very far from being unique. This does *not* happen in the presence of dynamical gravity. Quantum Gravity—more or less by definition—consists of several distinct topological sectors with very different geometries. So the *apparent* information loss in the IR with respect to UV in one sector is compensated by an information *gain* in a different sector; for instance, we get additional information in the IR from the winding states of a sector where one coordinate is compactified. However, we have a symmetric “information loss” in both IR and UV in the sense that in either limit, we see states which have one of the two quantum numbers m, w non-zero but not those which have both numbers non-zero.

Remark 6.3 We may rephrase our findings in a different way; the geometry of spacetime is not an *absolute* datum, but rather depends on the physical system we use to measure it, i.e. on the particular *probe*. We see one geometry by scattering with wave packets of momentum states and a different geometry using wave packets of winding states. In particular, the notion of “UV regime” is *not intrinsic* but refers to a particular choice of experimental set-up.

Remark 6.4 We have seen the surprising fact that as $R \rightarrow 0$ a new non-compact dimension “opens up”, which implies that a tower of *infinitely many* states (the winding modes) becomes *light*. This is the first example of a phenomenon which is believed to be *universal* in any consistent theory of Quantum Gravity. The theory has a non-compact moduli space of vacua, M , parametrized by the v.e.v. of light fields (here $\sigma \equiv \frac{1}{2} \log(\alpha' R^2)$), and as we go to *any* boundary at infinity of M , we should get an infinite tower of states whose masses go to zero exponentially [22, 25]. For $R \rightarrow \infty$, these light states are momentum modes, and for $R \rightarrow 0$, they are winding modes. In this example, the two infinite ends actually coincide, but in general M has several inequivalent infinite ends, each of them with its own tower of light modes.

T -Duality as a Spacetime Gauge Symmetry

It should be stressed that T -duality is a *duality* (i.e. a quantum equivalence) of the world-sheet QFT, and a *symmetry* of the spacetime physics which relates different states (backgrounds) of a single theory. In fact, as already claimed,¹⁷ in *closed oriented string theory*, T -duality is a spacetime **gauge** symmetry. That is, the compactifications at the two dual radii R and α'/R are not just two situations with the same physics, but actually *the same* physical configuration. Indeed, we saw around Eq. (6.113) that the linearized modulus $\delta\sigma \equiv \delta \log R$ at the self-dual radius is the $(3, 3)$ -component of a spacetime field transforming in the $(\mathbf{3}, \mathbf{3})$ of $SU(2) \times SU(2)$. The Weyl group of the first $SU(2)$, which is part of the Yang–Mills gauge symmetry, flips the sign of $\delta \log R$, thus implementing T -duality. In this way, T -duality gets identified with a gauge transformation in the $SU(2) \times SU(2)$ Yang–Mills group. This also implies that T -duality is an *exact symmetry* not just of string perturbation theory, but of the *exact* theory. For a more precise and general discussion of the relation between T -duality and the YM Weyl group, see the arithmetic part of Sect. 7.7.

T -Duality versus the Dilaton

The background value of the dilaton Φ varies under T -duality. The detailed path integral derivation of the dilaton transformation is given in Sect. 6.4.2; cf. Eq. (6.145). The conclusion is that the following combination is invariant under T -duality

$$e^{-2\Phi'} \sqrt{-G'} = e^{-2\Phi} \sqrt{-G} \quad (6.123)$$

This equality has a simple physical interpretation; consider the scattering of gravitons with momenta purely in the 25 non-compact dimensions. The amplitudes of physical processes should be independent of the chosen description, the original one or its T -dual. This is guaranteed by the equality (6.123). Indeed, in the compactification of one direction $y \sim y + 2\pi$ with $G_{yy} = R^2$ and $G'_{yy} = R'^2$, Eq. (6.123) becomes¹⁸

$$e^{-\Phi'} = e^{-\Phi} (R/R')^{1/2} = e^{-\Phi} (\alpha')^{-1/2} R \quad (6.124)$$

while the Einstein term in the 25d low-energy Lagrangian is (cf. Eq. (6.25))

$$\frac{\pi R}{2\kappa_0^2} e^{-2\Phi} \sqrt{-G} R \equiv \frac{\pi R'}{2\kappa_0^2} e^{-2\Phi'} \sqrt{-G'} R'. \quad (6.125)$$

¹⁷ This should be expected, since in any quantum system containing dynamical gravity—such as the string in critical dimension—all symmetries should be gauge symmetries [26].

¹⁸ As before, primed (unprimed) quantities refer to the T -dual (resp. original) description.

6.4.2 ★ T -Duality on a General Background and Buscher Rules

The above discussion refers to strings moving in the “trivial” KK geometry (6.1) with flat metric. In Sect. 6.1, we introduced more general KK geometries where M is the total space of a $U(1)$ principal bundle $M \rightarrow B$ and the massless background $\mathbf{G}_{MN}$, $\mathbf{B}_{MN}$, Φ is invariant under translation along the fibers. Moreover, in Sect. 6.4.1, we implicitly assumed that the world-sheet Σ is an infinite cylinder $\mathbb{C}^\times$ while, if the duality has to be consistent with string interactions, T -duality should be valid on arbitrary closed, oriented Σ 's. In this subsection, we address both generalizations.

The world-sheet action is (cf. Sect. 1.8)

$$\frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{h} \left(h^{\alpha\beta} \mathbf{G}_{MN} \partial_\alpha X^M \partial_\beta X^N + \epsilon^{\alpha\beta} \mathbf{B}_{MN} \partial_\alpha X^M \partial_\beta X^N + \alpha' \Phi R \right) \quad (6.126)$$

where $h_{\alpha\beta}$ is the world-sheet metric, R its scalar curvature, $\mathbf{G}_{MN}$, $\mathbf{B}_{MN}$, and Φ the background spacetime fields. We assume there is a Killing vector k which leaves invariant the background. More precisely, we assume k to act freely with closed orbits, so that our spacetime M is a $U(1)$ principal bundle $M \rightarrow B$ (cf. Sect. 6.1) while the $\mathbf{B}_{MN}$ and Φ backgrounds are $U(1)$ invariant: $\mathcal{L}_k \mathbf{B}_{MN} = \mathcal{L}_k \Phi = 0$. We may find a periodic angular coordinate y such that $k = \partial_y$. We write $2\pi R$ for its period. The action becomes

$$\begin{aligned} \frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{h} & \left(G_{yy} \partial^\alpha y \partial_\alpha y + 2 G_{yi} \partial^\alpha y \partial_\alpha X^i + G_{ij} \partial^\alpha X^i \partial_\alpha X^j + \right. \\ & \left. + \epsilon^{\alpha\beta} B_{yi} \partial_\alpha y \partial_\beta X^i + \epsilon^{\alpha\beta} B_{ij} \partial_\alpha X^i \partial_\beta X^j + \alpha' \Phi R \right) \end{aligned} \quad (6.127)$$

where the couplings are y -independent and we set $G_{yy} \equiv e^{2\sigma}$, $G_{yi} \equiv e^{2\sigma} A_i$. Note that the path integral is still Gaussian in the field y despite the fact that the background depends non-trivially on the coordinates x^μ of the base B of the principal bundle. We rewrite the action in the form

$$\begin{aligned} \frac{1}{4\pi\alpha'} \int d^2\sigma \sqrt{h} & \left(G_{yy} \eta^\alpha \eta_\alpha + 2 G_{yi} \eta^\alpha \partial_\alpha X^i + G_{ij} \partial^\alpha X^i \partial_\alpha X^j + \right. \\ & \left. + \epsilon^{\alpha\beta} B_{yi} \eta_\alpha \partial_\beta X^i + \epsilon^{\alpha\beta} B_{ij} \partial_\alpha X^i \partial_\beta X^j + \alpha' \Phi R + 2i \tilde{y} \epsilon^{\alpha\beta} \partial_\alpha \eta_\beta \right) \end{aligned} \quad (6.128)$$

where $\eta \equiv \eta_\alpha d\sigma^\alpha$ is a 1-form field on the world-sheet Σ (of genus g) and $\tilde{y}$ a Lagrange multiplier enforcing the condition $d\eta = 0$ whose *local* solution is $\eta = dy$ for some function y : the net effect of integrating out $\tilde{y}$ is to replace η_α by $\partial_\alpha y$, getting back the original action. On the other hand, if we integrate first in the Gaussian field η_α , we get the T -dual form of the action, proving the equivalence of the two formulations. However, while Gaussian integrals are *exact* at the quantum level, this formal treatment is not precisely correct in two respects; we must pay attention to the global aspects¹⁹ and to the precise functional measure. We address the two topics in turn.

Global Aspects: Dirac Quantization of Fluxes

y is a periodic field of period $2\pi R$, and its value is well-defined only mod $2\pi R$. In a fancy language, we say that it is a “zero-form” gauge field [27]. Its “field strength” dy is a globally defined 1-form which is d -closed by not d -exact; rather it represents a class $[dy] \in 2\pi R H^1(M, \mathbb{Z})$, i.e. its “field strength” has quantized fluxes $\oint dy \in 2\pi R \mathbb{Z}$. We write the last term in the action (6.128) as

$$- \frac{i}{2\pi\alpha'} \int d\tilde{y} \wedge \eta \quad (6.129)$$

¹⁹ For a general discussion of global aspects of dualities, see Sect. 1.6.1 in [27].

where η is a globally defined 1-form and $d\tilde{y}$ a d -closed but not d -exact 1-form. On-shell η is equal to $d\tilde{y}$. Writing $d\tilde{y} = h + df$ with h harmonic and f a global function, the path integral becomes

$$\int [d(\text{fields})] \exp \left\{ -S \Big|_{\partial_\alpha y \rightarrow \eta_\alpha} + \frac{i}{2\pi\alpha'} \int (h \wedge \eta - f d\eta) \right\} \quad (6.130)$$

The integral over f makes η closed. The integral over all harmonic forms h would set the harmonic part of η to zero, which implies $\eta = d\tilde{y}$ for a global function $\tilde{y}$: this prescription would give back the original action with $\tilde{y}$ a *non-compact* scalar. To get a compact field $\tilde{y}$ of period $2\pi R$, we have to restrict the class $[h]$ to be integral up to overall normalization, i.e. quantize the flux $\oint d\tilde{y}$ of the field strength $d\tilde{y}$ of $\tilde{y}$ *à la* Dirac. We write

$$h = 2\pi\lambda \sum_i n_i \omega_i, \quad n_i \in \mathbb{Z}, \quad \omega_i \text{ a basis of } H^1(\Sigma, \mathbb{Z}) \quad (6.131)$$

where λ is a real coefficient to be determined below by the Dirac condition. The path integral over $\tilde{y}$ factorizes into an integral over the global scalar f and a sum over its quantized harmonic projection h ("zero-modes" or "fluxes"). The sum is

$$\sum_{n_i \in \mathbb{Z}^g} \exp \left[\frac{i\lambda}{\alpha'} n_i \int \omega_i \wedge \eta \right] = \prod_{i=1}^g \delta_{\mathbb{Z}} \left(\frac{\lambda}{2\pi\alpha'} \int \omega_i \wedge \eta \right) \quad (6.132)$$

where $\delta_{\mathbb{Z}}(x) = \sum_{n \in \mathbb{Z}} \delta(x - n)$ cf. **BOX 6.4**. The condition that, on-shell, $\eta = d\tilde{y}$ with $\tilde{y}$ a periodic of period $2\pi R$ reads

$$\int \omega_i \wedge \eta = 2\pi R m_i \quad m_i \in \mathbb{Z} \quad (6.133)$$

which is the constraint enforced by $\delta_{\mathbb{Z}}$ in (6.132) provided we choose $\lambda = \alpha'/R$. Then

$$[d\tilde{y}] = h \in \frac{2\pi\alpha'}{R} H^1(\Sigma, \mathbb{Z}), \quad (6.134)$$

that is, the dual scalar field $\tilde{y}$ is also periodic of dual period

$$2\pi R' \equiv 2\pi\alpha'/R. \quad (6.135)$$

Performing the Gaussian integral over the 1-form field η , we get the T -dual form of the action where the original periodic scalar y of radius R is replaced by the dual periodic scalar $\tilde{y}$ of radius R' . The rules of Gaussian integration give the dual background

$$\tilde{G}_{yy} = 1/G_{yy}, \quad \tilde{G}_{yi} = B_{yi}/G_{yy}, \quad \tilde{B}_{yi} = G_{yi}/G_{yy} \quad (6.136)$$

$$\tilde{G}_{ij} = G_{ij} - (G_{yi}G_{yj} - B_{yi}B_{yj})/G_{yy}, \quad (6.137)$$

$$\tilde{B}_{ij} = B_{ij} - (G_{yi}B_{yj} - B_{yi}G_{yj})/G_{yy}. \quad (6.138)$$

These formulae are known as the *Buscher rules* [28, 29]. Let us reformulate them in the bundle language. The T -dual of the principal bundle $M \rightarrow B$ is another principal bundle $\tilde{M} \rightarrow \tilde{B}$ on a non-isometric but diffeomorphic base $\tilde{B} \simeq B$. The gauge connection on these bundle are (cf. (6.3))

$$A_i \equiv G_{iy}/G_{yy} = \tilde{B}_{iy}, \quad \tilde{A}_i \equiv \tilde{G}_{iy}/\tilde{G}_{yy} = B_{iy}, \quad (6.139)$$

so T -duality interchanges the KK connection A_i with the B -field $U(1)$ connection B_{iy} . In terms of their respective electric charges, this is (compact momentum) $\leftrightarrow$ (winding number), as expected. We have the relations

$$\tilde{\sigma} = -\sigma, \quad \tilde{G}_{ij} - e^{2\tilde{\sigma}} \tilde{A}_i \tilde{A}_j = G_{ij} - e^{2\sigma} A_i A_j. \quad (6.140)$$

Functional Measure: the Dilaton Shift

We have been a little cavalier with the functional measure. The shift of the dilaton under T -duality, Eq. (6.124), arises from subtleties with the Gaussian functional measure. Our naive treatment would be correct if the Gaussian fields we integrated over had no zero-mode. To get the correct answer, we need to treat the zero-modes carefully. This is hard in general, and we make the special assumption that G_{yy} , G_{yi} , and B_{yi} are slowly varying functions on the base B , i.e. they are almost constant on the string length scale $\sqrt{\alpha'}$. This assumption allows us to neglect the gradients of the background fields to the leading order in the α' -expansion, i.e. at 2d one-loop. Moreover, we reparametrize the scalar field so that its period is the self-dual one $2\pi\sqrt{\alpha'}$ by replacing

$$G_{yy} \rightsquigarrow R^2 G_{yy}/\alpha'. \quad (6.141)$$

In the path integral for the original fields, with action (6.127), there is a single²⁰ zero-mode for y . The zero-mode is the center of mass position in the $U(1)$ fiber and is integrated over the fiber

$$Z = \int_B d\text{vol} \int [d\phi]' e^{-S(\phi)} \int_0^{2\pi\sqrt{\alpha'}} \sqrt{G_{yy}} dy. \quad (6.142)$$

In the formulation with action (6.128), where we integrate on η and $\tilde{y}$, we should get back the correct measure on the zero-mode of the dual scalar $\tilde{y}$ if we first integrate out η . We decompose the 1-form η as $\sum_i h_i \rho_i + \xi$ where ξ is the component of η orthogonal to all harmonic forms, $\{\rho_i\}$ is an orthonormal basis of harmonic 1-forms, and $h_i \in \mathbb{R}^{2g}$ parametrize the harmonic projection of η . The path integral factorizes in an integral over the zero-modes and one on ξ . We write only the zero-mode factor

$$\sqrt{G_{yy}} \int \prod_{i=1}^{2g} \frac{dh_i}{\sqrt{2\pi^2\alpha'}} \exp \left[-\frac{1}{2\pi\alpha'} G_{yy} \sum_{i=1}^{2g} h_i^2 \right] \quad (6.143)$$

One gets the zero-mode measure (here $\chi(\Sigma) \equiv 2 - 2g$ is the Euler characteristic)

$$\sqrt{G_{yy}} G_{yy}^{-g} d\tilde{y} = \sqrt{\tilde{G}_{yy}} G_{yy}^{1-g} d\tilde{y} \equiv \exp \left[\frac{1}{2} \chi(\Sigma) \log G_{yy} \right] \sqrt{\tilde{G}_{yy}} d\tilde{y} \quad (6.144)$$

In view of the Gauss–Bonnet theorem, the extra factor with respect to (6.142) may be absorbed in a shift of the dilaton

$$\tilde{\Phi} = \Phi - \frac{1}{2} \log G_{yy} \equiv \Phi - \frac{1}{4} \log(G_{yy}/\tilde{G}_{yy}), \quad (6.145)$$

which is the last Buscher rule [28, 29].

6.4.3 Compactification of Several Dimensions

We generalize the discussion to the compactification of k dimensions

$$X^m \cong X^m + 2\pi R, \quad 26 - k \leq m \leq 25. \quad (6.146)$$

$d = 26 - k$ is the number of non-compact directions, and our spacetime is

²⁰ We ignore the zero-modes of the other scalar fields since they and their measure remain invariant through our duality manipulations.

$$\text{space-time} = \mathbb{R}^{d-1,1} \times T^k, \quad T^k \equiv (S^1)^k. \quad (6.147)$$

More generally, we may consider a non-trivial $U(1)^k$ principal bundle $M \rightarrow B$ over a $(26 - k)$ -dimensional base B .

The periodicity of the coordinates is kept fixed in (6.146) as a choice of field parametrization; the actual geometry depends on the internal metric G_{mn} . Now we have also a non-trivial internal 2-form B_{mn} living in the compact fiber T^k . The total number of massless scalars in d dimensions from the fields G_{mn} , B_{mn} is k^2 . In addition, we have k KK gauge bosons $G_{\mu m}$ and k gauge fields from the 2-form, $B_{\mu m}$. We already wrote the low-energy effective theory in Sect. 6.1:

$$\begin{aligned} \frac{(2\pi R)^k}{2\kappa_0^2} \int d^d x \sqrt{-G^d} e^{-2\Phi^d} \Big[& R_d + 4\partial_\mu \Phi^d \partial^\mu \Phi^d - \\ & - \frac{1}{4} G^{mn} G^{pq} (\partial_\mu G_{mp} \partial^\mu G_{nq} + \partial_\mu B_{mp} \partial^\mu B_{nq}) - \\ & - \frac{1}{4} G_{mn} F_{\mu\nu}^m F^{n\mu\nu} - \frac{1}{4} G^{mn} H_{m\mu\nu} H_n^{\mu\nu} - \frac{1}{12} H_{\mu\nu\lambda} H^{\mu\nu\lambda} \Big] \end{aligned} \quad (6.148)$$

where

$$\Phi^d = \Phi - \frac{1}{4} \log \det G_{mn}. \quad (6.149)$$

The String Spectrum

The new element is the 2-form background B_{mn} . The corresponding coupling in the world-sheet Lagrangian is proportional to

$$B_{mn} \partial_a (\epsilon^{ab} X^m \partial_b X^n). \quad (6.150)$$

When B_{mn} is constant (6.150) is *locally* a total derivative: a constant B_{mn} has no *local* effect, so the world-sheet theory is still a CFT, that is, a valid string background. However, the field X^m is not globally univalued, and (6.150) is a closed but not exact 2-form on the world-sheet. Hence, the coupling (6.150) does modify the physical spectrum. We work in the canonical quantization focusing on the zero-mode contribution (the only aspect affected by the coupling (6.150)). We write y and t for the world-sheet coordinates. The zero-mode part of the compact scalar fields is

$$X^m|_{\text{zero-mode}} = x^m(t) + w R y, \quad (6.151)$$

and the world-sheet action reduces to the Euclidean Lagrangian for $x^m(t)$

$$L = \frac{1}{2\alpha'} G_{mn} (\dot{x}^m \dot{x}^n + w^m w^n R^2) + \frac{i}{\alpha'} B_{mn} \dot{x}^m w^n R. \quad (6.152)$$

The canonical momenta are (we set $v^m \equiv i \dot{x}^m$)

$$p_m = -\frac{\partial L}{\partial \dot{x}^m} = \frac{1}{\alpha'} (G_{mn} v^n - B_{mn} w^n R) \quad (6.153)$$

The periodicity of the wave-functions implies quantization of the canonical momenta

$$p_m = \frac{n_m}{R} \Rightarrow v^m = \alpha' \frac{n_m}{R} + B_{mn} w^n R. \quad (6.154)$$

The zero-mode contribution to the world-sheet Hamiltonian is then

$$H|_{\text{zero-mode}} = \frac{1}{2\alpha'} G_{mn} (v^m v^n + w^m w^n R^2), \quad (6.155)$$

and the closed string mass is

$$m^2 = \frac{1}{2\alpha'^2} G_{mn} (v_L^m v_L^n + v_R^m v_R^n) + \frac{2}{\alpha'} (N + \tilde{N} - 2), \quad (6.156)$$

where

$$v_{L,R}^m = v^m \pm w^m R. \quad (6.157)$$

The constant background B_{mn} thus shifts the masses of the winding states through the dependence of v^m on B_{mn} , see (6.154). The $L_0 - \tilde{L}_0 = 0$ constraint becomes

$$0 = G_{mn} (v_L^m v_L^n - v_R^m v_R^n) + 4\alpha' (N - \tilde{N}) = 4\alpha' (n_m w^m + N - \tilde{N}), \quad (6.158)$$

which generalizes Eq. (6.98).

Partition Functions

We consider the torus path integral with the target space (6.147) and constant background fields G_{mn} and B_{mn} .

The B -field coupling is the integral of a closed 2-form on the world-sheet,

$$\frac{1}{4\pi} \int_{\Sigma} X^* B, \quad \text{where } B \equiv B_{mn} dX^m \wedge dX^n, \quad (6.159)$$

so it depends only on the topology ($\equiv$ homotopy class) of the field configuration

$$X: \Sigma \rightarrow \mathbb{R}^{d-1,1} \times T^k. \quad (6.160)$$

We specialize to $\Sigma = T^2$ and consider a configuration in the homotopy class of

$$X^m = (w_1^m \sigma^1 + w_2^m \sigma^2) R, \quad (w_1^m, w_2^m) \in \mathbb{Z}^{2k}, \quad (6.161)$$

where σ_1, σ_2 are angular coordinates in $T^2 \equiv (S^1)^2$. The “spatial” (resp. “time”) circle of the torus winds w_1^m (resp. w_2^m) times around the m th circle in target space. The value of the action B -term (6.159) on configurations in this topology class is

$$2\pi i b_{mn} w_1^m w_2^n \quad \text{where} \quad b_{mn} = B_{mn} R^2 / \alpha', \quad (6.162)$$

and the path integral takes the form

$$Z(\tau) = \sum_{(w_1, w_2) \in \mathbb{Z}^2} e^{2\pi i b_{mn} w_1^m w_2^n} Z_{w_1, w_2}(\tau) \quad (6.163)$$

with $Z_{w_1, w_2}(\tau)$ the path integral in the sector (w_1, w_2) in absence of B -field.

Exercise 6.5 Show that the partition function with the topological term (6.159) does reproduce, via Poisson resummation, the spectrum in Eqs. (6.156), (6.158).

“Flat-index” Fields It is convenient to introduce the internal vielbein e_m^r such that

$$G_{mn} = e_m^r e_n^r \quad (6.164)$$

with “flat”²¹ tangent indices $r, s, \dots$. The “flat index” coordinates

$$X^r \stackrel{\text{def}}{=} e_m^r X^m \quad (6.165)$$

then have the standard OPEs. The vertex operator momenta in this notation are

$$k_{Lr} \equiv e_r^m \frac{v_m L}{\alpha'}, \quad k_{Rr} \equiv e_r^m \frac{v_m R}{\alpha'} \quad (6.166)$$

with e_r^m the *inverse* vielbein $e_r^m e_m^s = \delta_r^s$. The mass-shell conditions take the form

$$m^2 = \frac{1}{2} (k_{Lr} k_{Lr} + k_{Rr} k_{Rr}) + \frac{2}{\alpha'} (N + \tilde{N} - 2) \quad (6.167)$$

$$0 = \alpha' (k_{Lr} k_{Lr} - k_{Rr} k_{Rr}) + 4(N - \tilde{N}), \quad (6.168)$$

where the second equation is equivalent to the $L_0 - \tilde{L}_0 = 0$ constraint (6.158). We shall use the orthonormal coordinates X^r in most constructions (often implicitly).

6.5 Narain Compactifications

We describe the general toroidal compactification on $(S^1)^k$. Consider the winding state vertex operator²²

²¹ Our use of the term “flat index” here follows the bizarre jargon of General Relativity. Since the metric G_{MN} is constant, the indices $r, s, \dots$ are simply vector indices in an *orthonormal basis*. The “flat index” coordinates X^r are nothing else than the standard Cartesian coordinates in the universal cover $\mathbb{R}^k$ of T^k .

²² To simplify the notation, we omit the cocycle factor and the symbol of normal order: both are implicitly assumed throughout.

$$e^{ik_L X_L + ik_R X_R}. \quad (6.169)$$

For a given toroidal compactification, the spectrum of allowed momenta (k_{Lr}, k_{Rr}) form a lattice in the $2k$ -dimensional momentum space $\mathbb{R}^{2k}$, i.e. the allowed momenta consist of all integral linear combinations of $2k$ linearly independent basis vectors. It is convenient to use the dimensionless momenta

$$l_{Lr} = k_{Lr}(\alpha'/2)^{1/2}, \quad l_{Rr} = k_{Rr}(\alpha'/2)^{1/2}, \quad (6.170)$$

and call the lattice where they take value Γ

$$(l_{Lr}, l_{Rr}) \in \Gamma \subset \mathbb{R}^{2k}. \quad (6.171)$$

The OPE of two vertices (6.169) reads

$$\begin{aligned} & : e^{ik_L X_L(z) + ik_R X_R(\bar{z})} : : e^{ik'_L X_L(w) + ik'_R X_R(\bar{w})} : \sim \\ & \sim (z-w)^{l_L l'_L} (\bar{z}-\bar{w})^{l_R l'_R} : e^{i(k_L + k'_L) X_L(w) + i(k_R + k'_R) X_R(\bar{w})} : \end{aligned} \quad (6.172)$$

As we carry one vertex around the other, the OPE picks up a phase

$$\exp\left[2\pi i(l_L \cdot l'_L - l_R \cdot l'_R)\right]. \quad (6.173)$$

Mutual locality of the various operators then requires

$$l \circ l' \stackrel{\text{def}}{=} l_L \cdot l'_L - l_R \cdot l'_R \in \mathbb{Z} \quad \text{for all } l, l' \in \Gamma. \quad (6.174)$$

The *circle* product $\circ$ has signature (k, k) in $\mathbb{R}^{2k}$. The dual lattice $\Gamma^\vee$ is the set of points in $\mathbb{R}^{2k}$ with integral $\circ$ -product with all vectors in Γ , that is,

$$\Gamma^\vee \stackrel{\text{def}}{=} \left\{ v \in \Gamma \otimes_{\mathbb{Z}} \mathbb{Q} : v \circ l \in \mathbb{Z} \quad \forall l \in \Gamma \right\}. \quad (6.175)$$

Thus *mutual locality* of the vertices is equivalent to $\Gamma \subset \Gamma^\vee$. We already know from Sect. 5.1 that modular invariance requires (in particular) Γ to be maximal with respect to locality, that is, to be *self-dual*, i.e.

$$\Gamma^\vee \equiv \Gamma. \quad (6.176)$$

Modular Invariance

Let us study modular invariance of a CFT with k compact scalars from scratch. Invariance under $T : \tau \mapsto \tau + 1$ requires $L_0 - \tilde{L}_0$ to be an integer for all states in the CFT. Since (cf. Eq. (6.168))

$$L_0 - \tilde{L}_0 = \frac{\alpha'}{4}(k_L \cdot k_L - k_R \cdot k_R) + N - \tilde{N} = \frac{1}{2}l \circ l + N - \tilde{N} \quad (6.177)$$

the condition is simply

$$l \circ l \in 2\mathbb{Z} \quad \text{for all } l \in \Gamma. \quad (6.178)$$

By definition, a lattice which satisfies this condition is called *even* (cf. **BOX 2.10**). Equation (6.178) implies (6.174) by polarization.²³

Invariance under $S: \tau \mapsto -1/\tau$ is more tricky, but from the analysis at the end of Sect. 5.1, we know that it is equivalent to the maximality condition (6.176). The partition function $Z(\tau) \equiv \text{Tr}[q^{L_0 - c/24} \bar{q}^{\tilde{L}_0 - c/24}]$ for k compact scalars is

$$Z(\tau)/(2\pi R)^k \equiv Z_\Gamma(\tau) = |\eta(\tau)|^{-2k} \sum_{l \in \Gamma} \exp[\pi i \tau l_L^2 - \pi i \bar{\tau} l_R^2] \quad (6.179)$$

since $L_0 = \frac{1}{2}l_L^2 + N$. We use Poisson resummation in the form²⁴ (here $x \in \mathbb{R}^{2k}$)

$$\delta_\Gamma(x) \stackrel{\text{def}}{=} \sum_{l \in \Gamma} \delta(x - l) = \frac{1}{\text{vol}(\Gamma)} \sum_{m \in \Gamma^\vee} \exp(2\pi i x \circ m) \quad (6.180)$$

where $\text{vol}(\Gamma)$ is the volume of a unit cell of Γ . Using this identity, we write

$$\begin{aligned} Z_\Gamma(\tau) &= \frac{1}{\text{vol}(\Gamma)} |\eta(\tau)|^{-2k} \sum_{m \in \Gamma^\vee} \int d^2l \exp[2\pi i m \circ l + \pi i \tau l_R^2 - i\pi \bar{\tau} l_L^2] = \\ &= \frac{1}{\text{vol}(\Gamma)} (\tau \bar{\tau})^{-k/2} |\eta(\tau)|^{-2k} \sum_{m \in \Gamma^\vee} \exp(-\pi i m_L^2/\tau + \pi i m_R^2/\bar{\tau}) = \\ &= \frac{1}{\text{vol}(\Gamma)} Z_{\Gamma^\vee}(-1/\tau) \end{aligned} \quad (6.181)$$

where in the last line we used the modular transformation of $\eta(\tau)$. As predicted, we conclude that the partition function is invariant under $\tau \rightarrow -1/\tau$ *iff* Γ is *self-dual* $\Gamma^\vee = \Gamma$, which also entails²⁵ $\text{vol}(\Gamma) = 1$. We summarize the result as follows

²³ If $q(v)$ is a quadratic form on a vector space V over a field $\mathbb{K}$ of characteristic $\neq 2$, we get a symmetric bilinear form $V \times V \rightarrow \mathbb{K}$ by polarizing it: $\langle v, w \rangle = \frac{1}{2}(q(v+w) - q(v) - q(w))$, for $v, w \in V$.

²⁴ *Proof* Let $\{e^i\} \in \mathbb{R}^{2k}$ be a set of generators of $\Gamma^\vee$, so that $\Gamma^\vee \ni m \equiv m_i e^i$ with $m_i \in \mathbb{Z}$. Then the sum in the RHS of (6.180) is $\sum_{m_i} \exp[2\pi i (x \circ e^i) m_i] = \prod_i \delta_{\mathbb{Z}}(x \circ e^i)$. The distribution in the RHS has support on Γ and is invariant by translation in Γ ; then it is $J^{-1} \delta_\Gamma(x)$ where the overall normalization constant J^{-1} is the inverse of the Jacobian $J = |\det e_a^i| = \text{vol}(\Gamma^\vee) = \text{vol}(\Gamma)^{-1}$.

²⁵ Note that $\text{vol}(\Gamma) \cdot \text{vol}(\Gamma^\vee) = 1$.

$Z_\Gamma(\tau)$ is modular-invariant *iff* Γ is a self-dual even lattice of signature (k, k)

As already mentioned, such lattices have been fully classified [30, 31]. Modular invariance depends on the lattice of allowed momenta (l_L, l_R) only through the *indefinite* product $\circ$ which is invariant under $O(k, k, \mathbb{R})$ rotations of the ambient space $\mathbb{R}^{2k}$. Therefore, if Γ is an even self-dual lattice of signature (k, k) , so are all lattices of the form

$$\Gamma' = \Lambda \Gamma \subset \mathbb{R}^{2k}, \quad \Lambda \in O(k, k, \mathbb{R}). \quad (6.182)$$

More generally, the space of even, self-dual lattices of *any* signature (r, s) organizes itself into complete orbits of the group $O(r, s, \mathbb{R})$.

We stress that the Lie group $O(k, k, \mathbb{R})$ is *not* a symmetry of the theory: the mass-shell conditions and OPEs involve separately the two dot products $l_L \cdot l'_L$ and $l_R \cdot l'_R$ which are preserved only by the *maximal compact* subgroup

$$O(k, \mathbb{R}) \times O(k, \mathbb{R}) \subset O(k, k, \mathbb{R}) \quad (6.183)$$

which acts independently on the left- and on the right movers. We illustrate the point in the $k = 1$ case, where (setting $r \equiv R(2/\alpha')^{1/2}$)

$$l_{L,R} = \frac{n}{r} \pm \frac{mr}{2} \quad n, m \in \mathbb{Z} \quad (6.184)$$

which form an even self-dual lattice; indeed

$$l \circ l' = mn' + nm' \in \mathbb{Z}. \quad (6.185)$$

The boost of “rapidity” η ,

$$l'_L = l_L \cosh \eta + l_R \sinh \eta, \quad l'_R = l_L \sinh \eta + l_R \cosh \eta, \quad (6.186)$$

changes the compactification radius as

$$r \rightarrow r' = r e^{-\eta}, \quad (6.187)$$

and hence modifies the string mass spectrum. The classification of even self-dual lattices will be studied in detail in Chap. 7. Here we quote the fundamental result:

Theorem 6.1 (e.g. [30, 31]) *Even self-dual lattices of signature (r, s) exists if and only if $r - s = 0 \bmod 8$. When this condition is satisfied, ALL even self-dual lattices of indefinite signature (r, s) make a single orbit under the group $O(r, s, \mathbb{R})$.*

In our case the even, self-dual lattices of signature (k, k) are all obtained from any given one Γ_0 by a $O(k, k, \mathbb{R})$ rotation

$$\begin{array}{c} \text{even, self-dual lattices} \\ \text{of signature } (k, k) \end{array} = \left\{ \Lambda \Gamma_0 : \Lambda \in O(k, k, \mathbb{R}) \right\}. \quad (6.188)$$

We can start, say, from the compactification with $B_{mn} = 0$ and all radii at the self-dual $SU(2)$ point. This lattice is usually written as

$$H^{\oplus k} \equiv \overbrace{H \oplus H \oplus \cdots \oplus H}^{k \text{ summands}} \quad (6.189)$$

since the matrix of the $\circ$ product is the direct sum of k copies of the hyperbolic pseudo-metric on $\mathbb{R}^2$, given by the Pauli matrix σ_1 , i.e.

$$\mathbf{v} \circ \mathbf{w} \equiv (v_1, v_2) \sigma_1 (w_1, w_2)^t \equiv v_1 w_2 + v_2 w_1. \quad (6.190)$$

Compactification at Γ_0 produces an enhanced gauge group $SU(2)^{2k}$; cf. Sect. 6.4.

Two $O(k, k, \mathbb{R})$ rotations, Λ and Λ' , yield physically equivalent configurations iff

$$\Lambda' \Lambda^{-1} \in O(k, \mathbb{R}) \times O(k, \mathbb{R}), \quad (6.191)$$

so that the inequivalent configurations are *locally* (in moduli space) parametrized by the Type III symmetric manifold (cf. **BOX 6.7**)

$$[O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R}) \equiv \widetilde{\mathcal{M}}. \quad (6.192)$$

The parametrization of (flat) toroidal backgrounds by points in the symmetric space $\widetilde{\mathcal{M}}$ is equivalent (but more convenient) to our previous description in terms of the background fields G_{mn} and B_{mn} . The dimension of the moduli space matches

$$\begin{aligned} \#G_{mn} + \#B_{mn} &= \frac{k(k+1)}{2} + \frac{k(k-1)}{2} = k^2 = \\ &= \frac{2k(2k-1)}{2} - 2 \frac{k(k-1)}{2} = \dim([O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R})) \end{aligned} \quad (6.193)$$

The equivalence of the two parametrizations is well-known in differential geometry.

Theorem 6.2 (Cartan [14]) *The Riemannian symmetric space*

$$[O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R}) \quad (6.194)$$

is globally diffeomorphic to the space of real $k \times k$ matrices whose symmetric part is positive definite, that is,

$$[O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R}) \simeq \{G_{mn} + B_{mn} \in \mathbb{R}(k) : G_{mn} > 0\}. \quad (6.195)$$

For the proof and additional details, see **BOX 6.8** where the diffeomorphism (6.195) between the two spaces is written explicitly.

Note 6.3 $O(k, k, \mathbb{R})$ is the $\mathbb{R}$ -split form of the complex Lie group $O(2k)$, see [32, 33]. More specifically, it is the Lie group of real-valued points in an algebraic group-scheme of Chevalley type [34].

6.5.1 The T -Duality Group

For $k > 1$, the T -duality group²⁶ is much richer than for $k = 1$. It is the subgroup of elements of $O(k, k, \mathbb{R})$ which map the reference lattice Γ_0 to itself. These rotations do not change the 2d CFT, just write it in a *different basis* for the lattice Γ_0 . The subgroup is written $O(k, k, \mathbb{Z})$ and consists of the $2k \times 2k$ matrices in $O(k, k, \mathbb{R})$ with integral entries.²⁷ It is an example of an *arithmetic group* [37, 38]. The two lattices $\Lambda\Gamma_0$ and $\Lambda\Lambda'\Gamma_0$, where $\Lambda' \in O(k, k, \mathbb{Z})$, are the same lattice, hence describe the *same* physical compactification of the string. We conclude

The arithmetic group $O(k, k, \mathbb{Z})$ is the Narain T -duality group

For $k = 1$, one gets back our old result: $O(1, 1, \mathbb{Z}) \simeq \mathbb{Z}_2 \times \mathbb{Z}_2$. The T -duality of (6.116) is given by

$$\pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \in O(1, 1, \mathbb{Z}). \quad (6.196)$$

In the general case, the space of physically inequivalent toroidal compactifications (called the *Narain moduli space*) is

$$\mathcal{M} \stackrel{\text{def}}{=} [O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R}) / O(k, k, \mathbb{Z}) \equiv \widetilde{\mathcal{M}} / O(k, k, \mathbb{Z}). \quad (6.197)$$

In terms of the action on the background fields G_{mn} , $B_{m,n}$ the T -duality group $O(k, k, \mathbb{Z})$ contains several different types of transformations. First, we have the T -dualities $R \rightarrow \alpha'/R$ along the various axes. Then we have the large spacetime coordinate transformations which respect the periodicity of coordinates, i.e.

$$x'^m = L^m_n x^n, \quad (6.198)$$

where L^m_n are integral matrices with $\det L = \pm 1$. The transformations (6.198) form the group $GL(k, \mathbb{Z})$, which is the group of homotopy classes of maps $T^k \rightarrow T^k$. Its index 2 subgroup $SL(k, \mathbb{Z}) \subset GL(k, \mathbb{Z})$ corresponds to maps which preserve the orientation. It is the *mapping class group of the k -torus T^k* . Finally, we have the integral shifts of the 2-form

²⁶ A general reference for this subsection is the review [21].

²⁷ For the *cognoscenti*: more intrinsically, $O(k, k, \mathbb{Z})$ is the *arithmetic group* of integer-valued points in the Chevalley group-scheme of type D_k associated to the lattice $(0) \oplus (v)$ of vector weights. For a review of Chevalley groups in the context of toroidal compactification of strings, see [34].

BOX 6.7 - Riemannian symmetric spaces of non-compact type

A Riemannian manifold M is *symmetric* if for *all* points $p \in M$ there is a involutive isometry $s_p: M \rightarrow M$, $s_p^2 = \text{Id}$, which fixes p and acts as -1 on the tangent space $T_p M$ at p . From this definition, it follows that a symmetric space is automatically (geodesically, hence metrically) *complete*. Alternatively, a Riemannian manifold M is (globally) *symmetric* if and only if it is simply connected, complete, and its Riemann tensor is *parallel* ($\equiv$ covariantly constant)

$$\nabla_i R_{jklm} = 0.$$

We are interested in the symmetric spaces of *non-compact* type, called Type III by Helgason [14]. The quotient G/K of a real semi-simple Lie group G by a maximal compact subgroup $K \subset G$, carries a (unique) structure of Riemannian symmetric space [14, 35, 36].

Identify the tangent space of the group manifold G with its Lie algebra $\mathfrak{g}$ via the Maurier–Cartan form $g^{-1}dg \in \Omega^1 \otimes \mathfrak{g}$. Let $\mathfrak{k} \subset \mathfrak{g}$ be the Lie algebra of the subgroup K and let

$$\mathfrak{p} = \mathfrak{k} \oplus \mathfrak{p}$$

be the (reductive) orthogonal decomposition with respect to the Killing form. Clearly, the tangent space of G/K gets identified with the vector space $\mathfrak{p}$ (for the *cognoscenti*, the *tangent bundle* to G/K is the G -homogeneous bundle canonically associated to the K -module $\mathfrak{p}$). Let $f: G/K \rightarrow G$ be any (locally defined) section of the K -principal bundle $G \rightarrow G/K$. $f^*(g^{-1}dg)$ is a 1-form on the coset G/K with coefficients in the Lie algebra $\mathfrak{g}$; take its projection

$$f^*(g^{-1}dg)_{\mathfrak{p}} \equiv e_i dx^i$$

on the subspace $\mathfrak{p}$ (we see the $e_i \in \mathfrak{g}$ as matrices *via* the adjoint representation). Then the G -invariant symmetric metric on G/K is

$$ds^2 = \text{tr}(e_i e_j) dx^i dx^j.$$

It is easy to check that ds^2 does not depend on the chosen local section f , so the metric is intrinsic and G -invariant (on the left). The other projection $f^*(g^{-1}dg)_{\mathfrak{k}}$ yields a K -connection on the principal bundle $G \rightarrow G/K$ which, being torsionless, should coincide with the Levi–Civita one. It is straightforward to check that the curvature tensor

$$R_{ijkl} = -\text{tr}([e_i, e_j][e_k, e_l])$$

is parallel and satisfies the negativity properties claimed in the main text.

Exercise 6.6 Show that the symmetric metric is Einstein, $R_{ij} = -\lambda G_{ij}$, for some $\lambda > 0$.

$$b_{mn} \rightarrow b_{mn} + N_{mn}, \quad N_{mn} = -N_{nm} \in \mathbb{Z}, \quad (6.199)$$

which leave invariant mod 2π the phase (6.162) hence the path integral (6.163). This is the 2d analogue of making $\theta \rightarrow \theta + 2\pi$ in the 4d Yang–Mills angle. These three kinds of transformations generate the full T -duality group $O(k, k, \mathbb{Z})$. From (6.199), we see that the d -dimensional massless scalar fields b_{mn} behave as *axions*.

More on the Narain Moduli Space

The moduli space $\mathcal{M}$ (6.197) is a *locally symmetric* space [14] of non-compact type, i.e. a quotient of the negative-curvature *symmetric* Riemannian manifold

BOX 6.8 - Geometry of the symmetric space $O(k, k)/[O(k) \times O(k)]$

We identify $O(k, k)$ with the group of $2k \times 2k$ matrices $\mathcal{E}$ such that

$$\mathcal{E}^t \Sigma \mathcal{E} = \Sigma, \quad \text{where} \quad \Sigma \equiv \left(\begin{array}{c|c} 0 & \mathbf{1}_{k \times k} \\ \hline \mathbf{1}_{k \times k} & 0 \end{array} \right)$$

$O(k) \times O(k)$ is the centralizer of Σ in $O(k, k)$ i.e. the group of $2k \times 2k$ matrices of the form

$$\left(\begin{array}{c|c} A & B \\ \hline B & A \end{array} \right) \quad \text{with} \quad (A \pm B)^t (A \pm B) = 1$$

Lemma (1) the real $k \times k$ matrix $\tau \equiv (A + B)(C + D)^{-1}$ has positive definite symmetric part. (2) all real $k \times k$ matrices with positive-definite symmetric part are of this form for some $\mathcal{E} \in O(k, k)$. (3) two elements $\mathcal{E}$ and $\mathcal{E}'$ of $O(k, k)$ yield the same matrix τ IFF $\mathcal{E}' = \mathcal{E}U$ with $U \in O(k) \times O(k)$. That is the space of symmetric, real, $k \times k$ matrices with positive-definite symmetric part is diffeomorphic to the Riemannian symmetric space $O(k, k)/[O(k) \times O(k)]$ via the above map.

Proof (1) exercise. (2) a real matrix τ with positive-definite symmetric part may be written uniquely in the form $\tau \equiv M + TT^t$ where $M^t = -M$ and T is a triangular matrix with positive diagonal elements; then the element

$$\mathcal{E} \equiv \left(\begin{array}{c|c} T & MT^{-t} \\ \hline 0 & T^{-t} \end{array} \right) \in O(k, k) \quad (\star\star)$$

does the job. (3) elementary computation. By **Iwasawa theorem**, all elements of $O(k, k)$ may be written—in a unique way—as a product of a matrix $(\star\star)$ and an element of $O(k) \times O(k)$.

The group $O(k, k)$ acts *transitively* on the space of real matrices with positive-definite symmetric part by multiplication on the *left* of the corresponding matrix $\mathcal{E}$ or, equivalently, by the map

$$\tau \mapsto (A\tau + B)(C\tau + D)^{-1} \quad \text{under the } O(k, k) \text{ rotation } \gamma = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array} \right) \in O(k, k)$$

Relation to String Backgrounds The identification of a given background configuration, G_{mn} , B_{mn} , of a string toroidal compactification with a point in the symmetric manifold $O(k, k)/[O(k) \times O(k)]$ is given by $B_{mn} + G_{mn} \stackrel{\text{def}}{=} \tau_{mn} \equiv M_{mn} + (TT^t)_{mn}$ the matrix T^t is the (triangular) vielbein of the internal metric G_{mn} and $M_{mn} \equiv B_{mn}$

Invariant Metric The unique (up to normalization) $O(k, k)$ -invariant metric on the space $O(k, k)/[O(k) \times O(k)]$ is constructed by the method in BOX 4.5:

$$\begin{aligned} ds^2 &\equiv \frac{1}{2} \text{tr} \left(\mathcal{E}^{-1} d\mathcal{E} + (d\mathcal{E}^t) \mathcal{E}^{-t} \right) = \\ &= \text{tr} \left(d(TT^t)(TT^t)^{-1} d(TT^t)(TT^t)^{-1} \right) + \text{tr} \left((dM)(TT^t)^{-1} dM^t (TT^t)^{-1} \right) = \\ &= G^{mn} (dG_{mp} + dB_{mp}) G^{pq} (dG_{nq} + dB_{nq}) \end{aligned}$$

$$\widetilde{\mathcal{M}} \equiv [O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R}) \quad (6.200)$$

by a discrete subgroup $\Gamma \equiv O(k, k, \mathbb{Z})$ of isometries acting properly discontinuously (but non-freely²⁸)

$$\mathcal{M} = \widetilde{\mathcal{M}}/\Gamma. \quad (6.201)$$

$\mathcal{M}$ is an *arithmetic quotient* (since the T -duality group is arithmetic). The arithmetic nature of the moduli space $\mathcal{M}$ introduces some additional deep structure and property [37, 38, 40].

The covering space $\widetilde{\mathcal{M}}$ with its unique (up to overall normalization) symmetric metric is a *Hadamard space*, i.e. a simply-connected, complete, Riemannian manifold with non-negative sectional curvatures.²⁹ By the Cartan–Hadamard theorem [35], $\widetilde{\mathcal{M}}$ is diffeomorphic to $\mathbb{R}^{k^2}$, and any two points in $\widetilde{\mathcal{M}}$ are connected by a *unique* geodesic. In particular, topologically, the moduli space $\mathcal{M}$ is a $K(\pi, 1)$ -space [8, 15] with $\pi \equiv O(k, k, \mathbb{Z})/\text{Tor}$, where the *finite* group $\text{Tor} \triangleleft O(k, k, \mathbb{Z})$ is the *torsion subgroup*.³⁰ Moreover, all non-trivial *finite-order* isometry of $\widetilde{\mathcal{M}}$ has a fixed point in $\widetilde{\mathcal{M}}$ [35]: the set of fixed points form a non-empty, convex, totally geodesic submanifold of dimension $< \dim \widetilde{\mathcal{M}}$. Conversely, an isometry which has *infinite* order has no fixed point. The same holds for finite *groups* of isometries; they have a non-empty fixed set.

In particular, the generic point in $\widetilde{\mathcal{M}}$ is not fixed by any non-trivial element of Γ . The special points which are fixed by some non-trivial subgroup $S \subset \Gamma$ correspond to the points in the moduli space $\mathcal{M}$ (“orbifold points”) where we have an *enhancement of gauge symmetry*. We have already seen this phenomenon in the case $k = 1$, see **BOX 6.9** for the rephrasing in the present “abstract” language.

Metric on the Narain Moduli Space

The moduli space is parametrized by light scalar fields ϕ^i ; their kinetic term

$$-\frac{1}{2} g_{ij}(\phi) \partial_\mu \phi^i \partial^\mu \phi^j \quad (6.202)$$

defines a *natural metric* $g_{ij}(\phi)$ on the moduli space $\mathcal{M}$. We stress that to get the proper definition of the kinetic-term metric, one has first to eliminate all mixing between the scalar fields’ and graviton kinetic terms; this is done by performing a spacetime Weyl transformation which makes the gravitational term $-\sqrt{-G}R$ independent of the scalar fields. Such canonical field parametrization is the *Einstein frame* (cf. Sect. 1.8.1). For the low-energy action (6.148), the Einstein frame scalars’ kinetic terms are

$$\frac{16}{d-2} \partial_\mu \Phi \partial^\mu \Phi + G^{mn} G^{pq} (\partial_\mu G_{mp} \partial^\mu G_{nq} + \partial_\mu B_{mp} \partial^\mu B_{nq}) \quad (6.203)$$

²⁸ By Minkowski theorem [37–39] (see footnote 9 in Chap. 5) Γ contains a finite index, normal subgroup Γ' which is torsionless and even neat. Hence we have a smooth *finite* cover of $\mathcal{M} \equiv \widetilde{\mathcal{M}}/\Gamma$ given by $\mathcal{M}' = \widetilde{\mathcal{M}}/\Gamma'$.

²⁹ In fact, it satisfies the *stronger* condition of having non-positive curvature operators.

³⁰ The torsion subgroup of any discrete group is the *normal* subgroup of its elements of finite order. Tor is obviously a *finite* group since it is both compact and discrete.

BOX 6.9 - T -duality revisited

For $k = 1$, we have

$$O(1, 1, \mathbb{R}) = \begin{pmatrix} \epsilon_1 \cosh x & \epsilon_1 \sinh x \\ \epsilon_2 \sinh x & \epsilon_2 \cosh x \end{pmatrix}, \quad x \in \mathbb{R}, \quad \epsilon_1, \epsilon_2 = \pm 1.$$

and we may identify the coset $\widetilde{\mathcal{M}}_{k=1} \equiv [O(1, \mathbb{R}) \times O(1, \mathbb{R})] \backslash O(1, 1, \mathbb{R})$ with the matrices of the above form with $\epsilon_1 = \epsilon_2 = 1$, that is, $\mathcal{M}_{k=1} \cong \mathbb{R}$ where we identify $x \equiv \log(R/\sqrt{\alpha'})$. Then the $k = 1$ T -duality group

$$O(1, 1, \mathbb{Z}) \equiv O(1, 1, \mathbb{R}) \cap \mathbb{Z}(2) = \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix}$$

acts on $\widetilde{\mathcal{M}}_{k=1} \cong \mathbb{R}$ as

$$\begin{aligned} \begin{pmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{pmatrix} &\rightarrow \begin{pmatrix} \cosh x & \sinh x \\ \sinh x & \cosh x \end{pmatrix} \begin{pmatrix} \epsilon_1 & 0 \\ 0 & \epsilon_2 \end{pmatrix} = \begin{pmatrix} \epsilon_1 \cosh x & \epsilon_2 \sinh x \\ \epsilon_1 \sinh x & \epsilon_2 \cosh x \end{pmatrix} \simeq \\ &\simeq \begin{pmatrix} \cosh x & \epsilon_1^{-1} \epsilon_2 \sinh x \\ \epsilon_1 \epsilon_2^{-1} \sinh x & \cosh x \end{pmatrix} \equiv \begin{pmatrix} \cosh(\pm x) & \sinh(\pm x) \\ \sinh(\pm x) & \cosh(\pm x) \end{pmatrix} \end{aligned}$$

i.e. T acts as $x \leftrightarrow -x$ or $R \leftrightarrow \alpha'/R$. The fixed point is the self-dual radius, where we have a gauge symmetry enhancement $U(1) \rightarrow SU(2)$. Identifying $\mathbb{R}$ with the Cartan subalgebra of $\mathfrak{su}(2)$, the T -duality group is nothing else than the Weyl group of $SU(2)$

$$\text{Weyl}(\mathfrak{su}(2)) \cong \mathbb{Z}_2.$$

This observation is valid in general: a point in moduli space where the gauge symmetry enhances from $U(1)^r$ to a non-Abelian group G of rank r corresponds to a point where the subgroup of the T -duality group fixing that point contains $\text{Weyl}(G)$.

The second term is precisely the $O(k, k, \mathbb{R})$ -invariant metric on the symmetric space $\widetilde{M}$, see BOX 6.8 for a detailed proof.

The group $O(k, k, \mathbb{R})$ is not a symmetry of the full string theory; only its discrete T -duality subgroup $O(k, k, \mathbb{Z})$ is. The difference between the two groups is not visible at low energy since it comes from the quantization of massive string modes. Thus $O(k, k, \mathbb{R})$ is an *accidental* symmetry of the low-energy physics. Intuitively this accidental symmetry looks “quite good”. Since $O(k, k, \mathbb{Z}) \subset O(k, k, \mathbb{R})$ is a (maximal) arithmetic subgroup, we may give a precise meaning to this idea. We write V for the vector representation of $O(k, k, \mathbb{R})$, $V^\vee$ for its dual, $T^{k,l} \equiv (\otimes^k V^\vee) \otimes (\otimes^l V)$ for the vector space of tensors with k covariant and l contravariant indices, and set $T^{\bullet,\bullet} = \oplus_{k,l} T^{k,l}$.

Theorem 6.3 (Borel density theorem [38, 41]) *If a tensor $t \in T^{\bullet,\bullet}$ is invariant under the T -duality group $O(k, k, \mathbb{Z})$, it is invariant for the full $O(k, k, \mathbb{R})$, that is, no algebraic-invariant may distinguish the two groups.*

In other words, no field-theoretic order parameter distinguishes the two groups. To detect the difference between $O(k, k, \mathbb{R})$ and $O(k, k, \mathbb{Z})$, we should look to subtler

physical observables without QFT counterparts which are sensitive to finer Number-Theoretic structures; the prime example is the quantum entropy of Black Holes.

We close this subsection with a further **Claim** which holds for all arithmetic quotients of symmetric spaces of non-compact type, i.e. double cosets of the form $K \backslash G(\mathbb{R}) / G(\mathbb{Z})$ where $G(\mathbb{Z}) \subset G(\mathbb{R})$ is an arithmetic subgroup (see **BOX 6.10**):

Claim 6.1 *For $k > 1$, $\mathcal{M} \equiv [O(k, \mathbb{R}) \times O(k, \mathbb{R})] \backslash O(k, k, \mathbb{R}) / O(k, k, \mathbb{Z})$ is a complete, non-compact, locally symmetric Riemannian manifold of finite total volume.*

6.6 Abelian Orbifolds

We may identify the scalar field X under the spacetime reflection

$$r: X^{25} \mapsto -X^{25}, \quad (6.204)$$

which amounts to gauging the $\mathbb{Z}_2$ symmetry group generated by the reflection. The fundamental domain is the half-line $X^{25} \geq 0$ with the hyperplane of fixed points

$$X^{25} = 0 \quad \textbf{fixed point locus} \quad (6.205)$$

as its boundary. More generally, we may identify spacetime under the *simultaneous* reflection of k coordinates

$$X^m \rightarrow -X^m \quad 26 - k \leq m \leq 25. \quad (6.206)$$

The locus of fixed points is now

$$X^{26-k} = \dots = X^{25} = 0. \quad (6.207)$$

For $k \geq 2$, the quotient of spacetime by (6.206) is singular; e.g. for $k = 2$, it is a conical singularity with angular deficit π .

We can combine the reflection (6.204) with the identification under the translation

$$t: X^{25} \rightarrow X^{25} + 2\pi R \quad (6.208)$$

to get a compact space. Identification of $\mathbb{R}$ under t yields a circle S^1 whose points we then identify under the action of r . The resulting compact space is a segment

$$0 \leq X^{25} \leq \pi R \quad (6.209)$$

as a fundamental region. The endpoints of the segments are the fixed points of r in S^1 . In other words, we are identifying the line $\mathbb{R}$ under the solvable group

BOX 6.10 - Volume of arithmetic quotients & Narain moduli spaces

We consider the following situation $G(\mathbb{R})$ is the Lie group of real-valued points of a simple Chevalley group (seen as an algebraic group scheme). In practice, $G(\mathbb{R})$ is the *split real form* of a simple Lie group. Then $K \subset G(\mathbb{R})$ is a maximal compact Lie subgroup, and $G(\mathbb{Z}) \subset G(\mathbb{R})$ is the maximal arithmetic subgroup of its integral points.

We want to compute the *finite* volume of the double coset $G(\mathbb{Z}) \backslash G(\mathbb{R}) / K$ where $G(\mathbb{R})$ is the split real form of a simple Lie group

$$\text{Vol}(G(\mathbb{Z}) \backslash G(\mathbb{R}) / K) = \text{Vol}(G(\mathbb{Z}) \backslash G(\mathbb{R})) / \text{Vol}(K).$$

The Narain module space has this form for

$$G(\mathbb{R}) = SO(k, k; \mathbb{R}) \quad \text{and} \quad K = SO(k) \times SO(k)$$

Theorem (Langlands [42]) *If $\zeta(\cdot)$ is the Riemann zeta function, $\prod_{i=1}^p (t^{2a_i-1} + 1)$ is the Poincaré polynomial of $G(\mathbb{C})$, and c is the order of the fundamental group of $G(\mathbb{C})$ then (for the standard normalization of the volume form, see [42])*

$$\text{Vol}(G(\mathbb{Z}) \backslash G(\mathbb{R})) = c \prod_{i=1}^p \zeta(a_i).$$

Remark By Hopf theorem, the $\{a_i\}$'s are the *degrees* of the fundamental Casimir invariants of $G(\mathbb{C})$, i.e. $a_i = \ell_i + 1$ with $\{\ell_i\}$ the *exponents* of $\mathfrak{g} \equiv \mathfrak{Lie}(G(\mathbb{C}))$ [43].

Corollary *The volume of the moduli group*

$$\mathcal{M}_d = G_d(\mathbb{Z}) \backslash G_d(\mathbb{R}) / K_d$$

is finite; indeed one has (for a standard normalization of the metrics)

$$\text{Vol}(\mathcal{M}_d) = \frac{c}{\text{Vol}(K_d)} \prod_{\ell=1}^r \zeta(d_\ell)$$

where $\{d_\ell\}$ are the degrees of the independent Casimir invariants of the real Lie group $G_d(\mathbb{R})$ and $\text{Vol}(K_d)$ is the volume of the compact Lie group K_d (computed by the Macdonald formula [44]); $\zeta(s)$ is the Riemann ζ -function.

Warning: one should pay attention to the relative normalization of the various volumes.

$$D_\infty \equiv \mathbb{Z} \rtimes \mathbb{Z}_2 \stackrel{\text{def}}{=} \langle t, r : r^2 = 1, rt = t^{-1}r \rangle \quad (6.210)$$

acting as (here $m \in \mathbb{Z}$)

$$t^m : X^{25} \cong X^{25} + 2\pi R m, \quad t^m r : 2\pi R m - X^{25}. \quad (6.211)$$

Similarly, we may identify the k -torus under the reflection (6.206). In this case, there are 2^k distinct fixed points where each coordinate X^m is either 0 or πR .

These singular spaces are examples of *orbifolds*, i.e. spaces which locally³¹ look like $\mathbb{R}^n$ quotiented by a finite group G which may not act freely. The singular set is the union of the fixed loci of elements $g \in G$, $g \neq 1$. A priori it is not obvious that string theory makes sense in such singular spacetimes, but it does [45, 46]. In fact a string theory is well-defined on all orbifolds not just the simple example we consider here:³² the main point is that the world-sheet 2d CFT is well-defined and modular-invariant since the geometric quotient by G can be implemented at the path integral level as the twisting procedure to be described in Sect. 6.6.1.

We return to the case of interest $\mathbb{R}/D_\infty$ where $G = \mathbb{Z}_2$. The identification of the target space under a reflection has two effects. First, the string wave-function should be invariant under the reflection, that is, it must be equal at identified points. Second, there is a new sector in the closed string spectrum satisfying the b.c.

$$X^{25}(\sigma^1 + 2\pi) = -X^{25}(\sigma^1) \quad (6.212)$$

since now the two sides represent the same point in the target space. Strings in this sector are called *twisted states*. The corresponding operators are called $\mathbb{Z}_2$ -twist fields. We shall see in Sect. 6.6.1 that modular invariance requires the twisted states to belong to the physical spectrum.

Strings on $\mathbb{R}^{24,1} \times S^1/\mathbb{Z}_2$

We focus on the compact one-dimensional orbifold $S^1/\mathbb{Z}_2$. In the untwisted sector, the spectrum of the theory on S^1 is reduced by the projection on the $\mathbb{Z}_2$ -invariant states. The effect of r on a general state is

$$|N, \tilde{N}; k^\mu, n, w\rangle \rightarrow (-1)^{k+\tilde{k}} |N, \tilde{N}; k^\mu, -n, -w\rangle, \quad (6.213)$$

where k ($\tilde{k}$) is the number of α_{-m}^{25} (resp. $\tilde{\alpha}_{-m}^{25}$) oscillators which act on the vacuum to produce the given state. In particular, from Eq. (6.213), we see that r *reverses* the compact winding number and momentum. The physical states are linear combinations invariant under this operation.

The states which are massless for generic values of the radius R have $n = w = 0$, so the projection simply requires that the total number of excitations of X^{25} , $k + \tilde{k}$, is *even*. The spacetime graviton, B -field, and dilaton survive the projection. The modulus of the compactification, associated to the radius R ,

$$\alpha_{-1}^{25} \tilde{\alpha}_{-1}^{25} |0; k_\mu, 0, 0\rangle \quad (6.214)$$

³¹ Most examples of orbifolds used in string theory are the so-called *perfect orbifolds*, that is, global quotients M/G of a smooth Riemannian manifold M by a discrete group G of isometries.

³² Here we consider only simple examples of orbifold where the local isotropy group is Abelian. There is a rich story for non-Abelian groups [45, 46] which we confine in the **Appendix**.

also survives, corresponding to the fact that R can take any value. However, the KK gauge bosons are projected out of the physical spectrum.

In the sector twisted by r , X^{25} is *anti-periodic* and hence has half-integral modes

$$X^{25}(z, \bar{z}) = i \left(\frac{\alpha'}{2} \right)^{1/2} \sum_{m=-\infty}^{+\infty} \frac{1}{m + 1/2} \left(\frac{\alpha_{m+1/2}^{25}}{z^{m+1/2}} + \frac{\tilde{\alpha}_{m+1/2}^{25}}{\bar{z}^{m+1/2}} \right) \quad (6.215)$$

The anti-periodicity forbids any center of mass coordinate or momentum, so *the string cannot move away from the $X^{25} = 0$ fixed point* (a part for quantum oscillations). There are also string states localized at the other fixed point

$$X^{25}(\sigma^1 + 2\pi) = 2\pi R - X^{25}(\sigma^1), \quad (6.216)$$

in this sector, the fields change by tr as we go along the closed string. The mode expansion is as in (6.215) plus a constant term πR . All other fixed points $n\pi R$ of D_∞ are images under $t^{\mathbb{Z}}$ of these two.

According to **BOX 1.2**, the zero-point energy of a free 2d scalar with periodic boundary conditions is $-1/24$ and with *anti-periodic* b.c. is $+1/48$: the difference is $+1/16$. The twist fields which generate the twist sector out of the vacuum $|0\rangle$ (which belongs to the untwisted sector) have Virasoro weights

$$(h, \tilde{h}) = \left(\frac{1}{16}, \frac{1}{16} \right). \quad (6.217)$$

The mass-shell condition for a twisted sector state of the bosonic string moving in $\mathbb{R}^{24,1} \times (S^1)/\mathbb{Z}_2$ is then

$$m^2 = \frac{4}{\alpha'} \left(N - \frac{15}{16} \right), \quad N = \tilde{N}. \quad (6.218)$$

Moreover, the oscillators of X^{25} give half-integral contributions to the level N . The r -projection again requires the total number of 25-excitations $k + \tilde{k}$ to be *even*. In the twisted sectors, the ground states are $|T_a\rangle$, where the index $a = 1, 2$ labels the two fixed points. These states are tachyonic, as are the first excited states

$$\alpha_{-1/2}^{25} \tilde{\alpha}_{-1/2}^{25} |T_a\rangle, \quad a = 1, 2. \quad (6.219)$$

There are no massless twisted states. The extension to $T^k/\mathbb{Z}_2$ is straightforward.

Correlation Functions of $\mathbb{Z}_2$ -Twists

Tree level amplitudes of *untwisted* states are easy; on the sphere, with all external states untwisted, the twists and projections do not enter into the calculation, and all the amplitudes are the same as in the untwisted theory. For example, the low-energy effective action for the massless untwisted fields $g_{\mu\nu}$, $B_{\mu\nu}$, Φ , $g_{25,25}$ is the same one we found in toroidal compactification, Eq. (6.148), with the vector fields omitted. This fact is convenient when studying the physics of an orbifold theory; it is called the *inheritance principle*.

The correlators of twisted state vertices are less easy to write explicitly. Amplitudes with an *odd* number of twisted-vertex insertions vanish. To compute the correlation function with $2g + 2$ twisted operators inserted at the points $p_i \in \Sigma$, we replace the world-sheet Σ by its double cover $\Sigma' \rightarrow \Sigma$ branched at the points p_i . Σ' exists by **Riemann's existence theorem** [47, 48]. The field X^{25} is single valued on Σ' and odd under the deck transformation of the cover. We compute the path integral on Σ' summing over odd configurations. This procedure is the obvious generalization of the well-known method of images. For instance, for sphere amplitudes with $\mathbb{Z}_2$ -twist operators inserted at points z_i , Σ' is the genus g *hyperelliptic curve* [49]

$$y^2 = \prod_{i=1}^{2g+2} (z - z_i), \quad (6.220)$$

see [50] for details.

The Partition Function of the $S^1/\mathbb{Z}_2$ CFT

In the untwisted sector, one has just to project on the $r = +1$ states

$$(q\bar{q})^{-1/24} \text{Tr}_{\text{untw.}} \left(\frac{1+r}{2} q^{L_0} \bar{q}^{\tilde{L}_0} \right) \quad (6.221)$$

The first term in the projector yields $1/2$ the partition function (6.48) for the toroidal compactification. The term with r inserted must have $n = w = 0$ as we see from Eq. (6.213). The factor $(-1)^{k+\tilde{k}}$ inserts signs in the oscillator sum with the effect

$$\prod_{m \geq 1} |1 - q^m|^{-2} \rightarrow \prod_{m \geq 1} |1 + q^m|^{-2}. \quad (6.222)$$

Thus the partition function of the *untwisted* sector is

$$Z(R, \tau)_{S^1/\mathbb{Z}_2, \text{untw.}} = \frac{1}{2} Z(R, \tau)_{S^1} + \frac{1}{2} (q\bar{q})^{-1/24} \prod_{m=1}^{\infty} |1 + q^m|^{-2}. \quad (6.223)$$

The contribution from the twisted sectors is a product on half-integral modes

$$\begin{aligned}
2 \times \frac{1}{2} (q\bar{q})^{1/48} \text{Tr}_{\text{twist}} \left(\frac{1+r}{2} q^{L_0} \bar{q}^{\tilde{L}_0} \right) &= \\
&= (q\bar{q})^{1/48} \left[\prod_{m \geq 1} |1 - q^{m-1/2}|^{-2} + \prod_{m \geq 1} |1 + q^{m-1/2}|^{-2} \right].
\end{aligned} \tag{6.224}$$

The extra factor 2 in (6.224) is the number of twisted sectors, i.e. of fixed points. Putting everything together, the orbifold partition function³³ is written as³⁴

$$\begin{aligned}
Z(R, \tau)_{S^1/\mathbb{Z}_2} &= \frac{1}{2} Z(R, \tau)_{S^1} + \left| \frac{\eta(\tau)}{\vartheta_{10}(\tau)} \right| + \left| \frac{\eta(\tau)}{\vartheta_{01}(\tau)} \right| + \left| \frac{\eta(\tau)}{\vartheta_{00}(\tau)} \right| \\
&= \frac{1}{2} \left(Z(R, \tau)_{S^1} + \left| \frac{\vartheta_{00} \vartheta_{01}}{\eta^2} \right| + \left| \frac{\vartheta_{00} \vartheta_{10}}{\eta^2} \right| + \left| \frac{\vartheta_{10} \vartheta_{01}}{\eta^2} \right| \right),
\end{aligned} \tag{6.225}$$

where in the second line we used the identity ($\diamond$) in **BOX 6.11** to get the standard expression one finds in the literature [51–53]. The first term in the RHS of the first line is already known to be modular-invariant, while we check from the known modular transformations of the η and ϑ -functions (**BOX 5.5**) that the sum of the other three terms is modular-invariant as well. In terms of path integrals on the torus, the term $Z(R, \tau)_{S^1}$ comes from fields which are periodic up to translations, while the term $|\eta/\vartheta_{\alpha\beta}|$ arises from fields configurations such that

$$X^{25}(z + 2\pi) = (-1)^{\alpha+1} X^{25}(z), \quad X^{25}(z + 2\pi\tau) = (-1)^{\beta+1} Z^{25}(z). \tag{6.226}$$

E.g. the ϑ_{10} term arises from the untwisted sector with r inserted in the trace, while ϑ_{01} arises from the trace over the twisted sectors. These two contributions are interchanged by S : $\tau \mapsto -1/\tau$. Hence we see that we should keep the twisted sector if the theory should be modular-invariant (i.e. consistent). Indeed, the algebra of r -even operators without twist fields is not maximal local.

6.6.1 Twisting Procedure

Both the S^1 compactification and the orbifold are particular instances of a general construction in string theory known as *twisting* (or *modding out*) by a symmetry. Starting from a given left-right symmetric CFT invariant under some discrete group H , we can form a new CFT by gauging the discrete symmetry. For simplicity, here we assume H to be Abelian, referring the reader to **Appendix 1** for the general case.

³³ Note that each term in the sum (6.225) is the *inverse* of the partition function for a complex fermion with the same boundary conditions on the torus. This reflects the fact that inverting the statistics from fermionic to bosonic inverts the free partition function.

³⁴ ϑ -functions notation as in **BOX 5.5**; cf. the product expressions for ϑ -functions in **BOX 6.11**.

We write $\hat{h}$ for the operator implementing the transformation $h \in H$ on the closed string Hilbert space. The H -twisting procedure goes through two steps.

BOX 6.11 - Some useful identities between ϑ -functions

From BOX 5.5

$$\begin{aligned}\vartheta_{10}(\tau) &= 2q^{1/8} \prod_{m \geq 1} (1 - q^m)(1 + q^m)^2 = \sum_{n \in \mathbb{Z}} q^{\frac{1}{2}(n+1/2)^2} \\ \vartheta_{00}(\tau) &= \prod_{m \geq 1} (1 - q^m)(1 + q^{m-1/2})^2 = \sum_{n \in \mathbb{Z}} q^{\frac{1}{2}n^2} \\ \vartheta_{01}(\tau) &= \prod_{m \geq 1} (1 - q^m)(1 - q^{m-1/2})^2 = \sum_{n \in \mathbb{Z}} (-1)^n q^{\frac{1}{2}n^2}\end{aligned}$$

From the product representations, we have

$$\vartheta_{10}(\tau) \vartheta_{00}(\tau) \vartheta_{01}(\tau) = 2\eta(\tau)^3, \quad (\diamond)$$

while from the series representations

$$\vartheta_{00}(\tau) + \vartheta_{10}(\tau) = \sum_{n \in \mathbb{Z}} q^{\frac{1}{8}n^2}, \quad \vartheta_{00}(\tau) + \vartheta_{0,1}(\tau) = 2 \sum_{n \in \mathbb{Z}} q^{2m^2}$$

multiplying these two identities and subtracting ϑ_{00}^2 yields

$$\vartheta_{00}(\tau) \vartheta_{10}(\tau) + \vartheta_{00}(\tau) \vartheta_{01}(\tau) + \vartheta_{01}(\tau) \vartheta_{10}(\tau) = 2 \sum_{n, w \in \mathbb{Z}} q^{2n^2 + \frac{1}{8}w^2} - \sum_{n, w \in \mathbb{Z}} q^{\frac{1}{2}(n^2 + w^2)} \quad (\clubsuit)$$

First step: we add to the Hilbert space the *twisted sectors*

$$\mathbf{H} = \bigoplus_{h \in H} \mathbf{H}_h. \quad (6.227)$$

Here $\mathbf{H}_h$ is the space of closed string states which satisfy the periodic b.c. twisted by the action of $h \in H$, i.e. for all local operators $\phi(w)$:

$$\phi(w + 2\pi) = \hat{h} \phi(w) \hat{h}^{-1} \equiv h \cdot \phi(w) \quad \text{in } \mathbf{H}_h. \quad (6.228)$$

Second step: we restrict the Hilbert space $\mathbf{H}$ to its H -invariant subspace. That is, after the H -twisting the physical Hilbert space is

$$\mathbf{H}_{\text{tw}} = P_H \mathbf{H} \quad \text{where} \quad P_H = \frac{1}{|H|} \sum_{h \in H} \hat{h}. \quad (6.229)$$

The vertex operator $\mathcal{O}_h(w)$ corresponding to a state $|\mathcal{O}_h\rangle_h \in \mathbf{H}_h$ produces branch cuts in the fields $\phi(w)$ with discontinuity $h \cdot \phi(w) - \phi(w)$, but the projection P_H onto h -invariant states/operators eliminates these branch cuts from the OPEs of vertex operators of states in $\mathbf{H}$. The OPEs close because the product of H -invariant operators is H -invariant. The projection is preserved by the string interactions since h is a symmetry. Therefore, if the Hilbert space $\mathbf{H}_{h=1}$ of the original CFT is linearly isomorphic to a local operator algebra $\mathfrak{A}_{h=1}$, the resulting H -twisted Hilbert space $\mathbf{H}_{\text{tw}}$ is also linearly isomorphic to a *local* algebra $\mathfrak{A}_H$. Moreover, if $\mathfrak{A}_{h=1}$ was maximal local so is $\mathfrak{A}_H$. Then *the twisted CFT is modular-invariant if and only if $h - \tilde{h} \in \mathbb{Z}$ for all elements of $\mathfrak{A}_H$.*

Note 6.4 The GSO projection is a special case of the twisting procedure where $H \equiv \mathbb{Z}_2$ acts by flipping the sign of fermions.

Partition Function The partition function of the H -twisted CFT

$$Z_{H\text{-tw}} = \frac{1}{|H|} \sum_{h_1, h_2 \in H} \text{Tr}_{\mathbf{H}_{h_1}} \left[\hat{h}_2 q^{L_0 - c/24} q^{\tilde{h} - c/24} \right] = \frac{1}{|H|} \sum_{h_1, h_2 \in H} Z^{h_1}_{h_2}(\tau) \quad (6.230)$$

is given by a sum of path integrals $Z^{h_1}_{h_2}(\tau)$ on the torus of periods $(2\pi, 2\pi\tau)$ where the fields are twisted along the spatial A -cycle and the time-like B -cycle by elements $h_1, h_2 \in H$, respectively. *Naïve* manipulations of the path integral will give

$$Z^{h_1}_{h_2}(\tau) \stackrel{?}{=} Z^{h_2}_{h_1^{-1}}(-1/\tau), \quad Z^{h_1}_{h_2}(\tau) \stackrel{?}{=} Z^{h_1}_{h_1 h_2}(\tau + 1), \quad (6.231)$$

and the sum (6.230) looks *naïvely* modular-invariant. However, the naive invariance may be spoiled by *anomalous phases* in the modular transformations (6.231); the ratio of the two sides is a phase which may be $\neq 1$, as we saw for the functions $Z^\alpha_\beta(\tau)$ in Sect. 5.4. For a right-left symmetric path integral, the anomalous phases cancel automatically, and the naive argument suffices.³⁵ There are no anomalous phases in the transformation $S: \tau \mapsto -1/\tau$, so the first Eq. (6.231) is correct.³⁶ The dangerous phases arise in the transformation $T: \tau \mapsto \tau + 1$, where they measure the failure of the matching condition $L_0 - \tilde{L}_0 \in \mathbb{Z}$.

In the special case of interest in this section, *left-right symmetric Abelian orbifolds*, where the same twisted b.c. (6.228) is imposed on both left- and right movers, all anomalous phases cancel, and the partition function is modular-invariant on the nose. We stress that if we had only inserted the projection on the H -invariant states, but not the twisted sectors, modular invariance would be trivially destroyed.

³⁵ This is the *diagonal* modular-invariant we already encountered in Remark 5.8.

³⁶ **Proof:** The ratio $Z^{h_1}_{h_2}(\tau)/Z^{h_2}_{h_1^{-1}}(-1/\tau)$ is both holomorphic and a phase, hence constant by Liouville theorem. Computing this constant at the fixed point of S , $\tau = i$, we conclude that the ratio is 1 without extra phases.

Discrete Torsion

When the integral spin condition $L_0 - \tilde{L}_0 \in \mathbb{Z}$ is satisfied, there may be *more than one* modular-invariant partition function, and hence more than one consistent string theory. The first example of this multiplicity are the *two* standard GSO $\pm$ projections which keep spacetime fermions of opposite chirality. This “old” example fits in the general theory of *discrete torsion* [54] to be introduced in this subsection.³⁷

Consider a CFT twisted by an Abelian symmetry group H with modular-invariant partition function (6.230). The CFT theory with partition function

$$Z(\tau) = \frac{1}{\text{order}(H)} \sum_{h_1, h_2 \in H} \varepsilon(h_1, h_2) Z_{h_2}^{h_1}(\tau), \quad (6.232)$$

is also *consistent*, i.e. modular-invariant with closed and local OPEs, provided the new extra phases $\varepsilon(h_1, h_2)$ satisfy the two properties:

$$(a) \text{ alternating:} \quad \varepsilon(h, h) = 1, \quad \varepsilon(h_1, h_2) = \varepsilon(h_2, h_1)^{-1} \quad (6.233)$$

$$(b) \text{ bilinear:} \quad \varepsilon(h_1, h_2) \varepsilon(h_1, h_3) = \varepsilon(h_1, h_2 h_3). \quad (6.234)$$

In terms of the operators $\hat{h}_2$ defined in the original twisted CFT, the new twisted theory is no longer projected into H -invariant states, but onto states which satisfy

$$\hat{h}_2 |\psi\rangle_{h_1} = \varepsilon(h_2, h_1) |\psi\rangle_{h_1} \quad (6.235)$$

in the sector twisted by h_1 . In other words, states are now eigenvectors of the operators $\hat{h}$ with a sector-dependent phase eigenvalue. Equivalently, we have made a sector-dependent redefinition

$$\hat{h} \rightarrow \varepsilon(h_1, h) \hat{h}. \quad (6.236)$$

The phase factor $\varepsilon(h_1, h_2)$ is known as the *discrete torsion* [54]. For the geometric interpretation of a string moving in an orbifold with discrete torsion, see [55].

Conditions (6.233) and (6.234) have an elegant interpretation in *group cohomology* [56]. The function $\varepsilon: H \times H \rightarrow U(1)$ is an element of $\text{Hom}(\wedge^2 H, U(1))$ and each such element is the image of a class $[f]$ in the second cohomology group $H^2(H, U(1))$ of the group H with coefficients in $U(1)$ (where H acts trivially on $U(1)$), by the map $f(h_1, h_2) \mapsto f(h_1, h_2) - f(h_2, h_1) \equiv \varepsilon(h_1, h_2)$ see **Exercise V.9.5** in [56].

Exercise 6.7 Prove that the OPE close and the partition function is modular precisely when $\varepsilon(h_1, h_2)$ satisfy the conditions (6.233), (6.234).

³⁷ Let H be the 2-torsion group generated by $(-1)^F$ and $(-1)^{\tilde{F}}$ where F and $\tilde{F}$ are the left- and right-moving fermion numbers of the closed superstring. Then taking the phase in Eq. (6.232) to satisfy (6.233), (6.234) with $\varepsilon((-1)^F, (-1)^{\tilde{F}}) = -1$ we interchange IIB $\leftrightarrow$ IIA.

Why “discrete torsion”? In a toroidal compactification on T^k the only effect of switching on a flat ($\equiv$ constant) background B_{ij} is to multiply the partition function Z_{m^i, n^j} over configurations in the homotopy class $(m^i, n^j) \in \mathbb{Z}^{2k}$ by a phase

$$\varepsilon: \mathbb{Z}^k \times \mathbb{Z}^k \rightarrow U(1), \quad \varepsilon(\mathbf{m}, \mathbf{n}) = \exp[2\pi i b_{ij} m^i n^j], \quad (6.237)$$

cf. Eq. (6.163). Clearly, $\varepsilon(\mathbf{m}, \mathbf{n})$ is a phase which satisfies (6.233), (6.234). Thus a discrete torsion is analogous to a flat 2-form background. A background with a non-trivial B -field is said to have *torsion*.³⁸

6.6.2 ★ More on the Kosterlitz–Thouless Transition Point

In the final paragraph of Sect. 6.2, we stated that the compact scalar at $R = \frac{1}{2}\sqrt{\alpha'}$ is the Kosterlitz–Thouless multi-critical CFT. This CFT has two *inequivalent* marginal deformations: the one obtained by varying R and a subtler one. We claim that the subtler CFT is a $S^1/\mathbb{Z}_2$ orbifold [17, 45, 50]; more precisely

Claim 6.2 *The compact scalar at $R = \frac{1}{2}\sqrt{\alpha'}$ is equivalent to the $S^1/\mathbb{Z}_2$ orbifold at $R = \sqrt{\alpha'}$.*

Proof We start from the compact scalar at the self-dual radius $\sqrt{\alpha'}$, i.e. from the $SU(2) \times SU(2)$ current algebra at level 1. We consider two $\mathbb{Z}_2$ subgroups of the symmetry group. First we have translations in S^1 by $\pi R \equiv \pi\sqrt{\alpha'}$ which is generated by $\exp[\pi i \hat{f}(j_3 + \tilde{j}_3)] \in SU(2) \times SU(2)$; second we have $X \rightarrow -X$ which is generated by (say) $\exp[\pi i \hat{f}(j_1 + \tilde{j}_1)] \in SU(2) \times SU(2)$. By construction, twisting with respect to the first $\mathbb{Z}_2$, we get the compact scalar at $R = \frac{1}{2}\sqrt{\alpha'}$, whereas twisting by the second one, we get the $S^1/\mathbb{Z}_2$ orbifold at $R = \sqrt{\alpha'}$. But the two $\mathbb{Z}_2$ ’s are conjugate in the symmetry group $SU(2) \times SU(2)$ and so produce equivalent CFTs. $\square$

Exercise 6.8 Show that all marginal deformation at the Kosterlitz–Thouless point leads to either a compact scalar CFT or a $S^1/\mathbb{Z}_2$ orbifold CFT.

Twist fields The twist fields of the orbifold at $R = \sqrt{\alpha'}$ should correspond to operators of the compact scalar at $R = \frac{1}{2}\sqrt{\alpha'}$. They are (we set $\alpha' = 2$)

$$\sigma_1 \equiv \sqrt{2} \cos \left[\frac{\sqrt{2}}{4} (X_L - X_R) \right], \quad \sigma_2 \equiv \sqrt{2} \sin \left[\frac{\sqrt{2}}{4} (X_L - X_R) \right] \quad (6.238)$$

which have $(h, \tilde{h}) = (\frac{1}{16}, \frac{1}{16})$ as they should.

Partition Functions From Eq. (6.225) the $S^1/\mathbb{Z}_2$ partition function at radius R has the form

$$Z(R)_{S^1/\mathbb{Z}_2} = \frac{1}{2} Z(R)_{S^1} + Z_{\text{twist}}, \quad (6.239)$$

where Z_{twist} is independent of R . We can compute it by specializing the formula to $R = \sqrt{2}$ where we know that $Z(\sqrt{2})_{S^1/\mathbb{Z}_2} = Z(\frac{1}{2}\sqrt{2})_{S^1}$ (in unit where $\alpha' = 2$). We get the remarkable identity

³⁸ This is a bit of abuse of language, since the actual *torsion* is the 3-form field strength $H \equiv dB$ which vanishes for B flat. However, the abusive language is by now standard.

$$Z(\tfrac{1}{2}\sqrt{2}; \tau)_{S^1} - \frac{1}{2}Z(\sqrt{2}; \tau)_{S^1} = Z(\tau)_{\text{twist}} = \left| \frac{\eta(\tau)}{\vartheta_{10}(\tau)} \right| + \left| \frac{\eta(\tau)}{\vartheta_{01}(\tau)} \right| + \left| \frac{\eta(\tau)}{\vartheta_{00}(\tau)} \right|. \quad (6.240)$$

Using the expression for $Z(R, \tau)_{S^1/\mathbb{Z}_2}$ in the second line of (6.225) and the formula (6.48) for $Z(R, \tau)_{S^1}$, we see that, for q real, Eq. (6.240) reduces to the identity (♣) in BOX 6.11. See also [51].

Even more remarkably, the partition functions of the scalar at $R = \sqrt{2}/2$ and the orbifold at $R = \sqrt{2}$ should be equal on surfaces of arbitrary genus g . However, for $g \geq 3$, they are not identical as functions of the period matrix τ_{ij} . In fact, the equality holds only on a subspace $\mathcal{S}_g \subset \mathcal{H}_g$ of the Siegel upper half-space; this is the Schottky locus of the periods of actual Riemann surfaces [17]. Thus we can determine the Schottky locus using the duality between the orbifold CFT on $S^1/\mathbb{Z}_2$ at $R = \sqrt{\alpha'}$ and the compact scalar at $\frac{1}{2}\sqrt{\alpha'}$.

Note 6.5 For the explicit expressions of the partition and correlation functions for both the compact scalar and the $S^1/\mathbb{Z}_2$ orbifold at arbitrary genus g , see [17, 50].

6.7 Open Strings: Adding Wilson Lines

The new feature in toroidal compactifications of open strings is the possibility of non-trivial *Wilson lines*, i.e. non-trivial flat backgrounds for the gauge connection associated to the Chan–Paton d.o.f. We start by reviewing the situation in QFT.

The Point-Particle Case For the sake of comparison, we preliminary consider a particle with electric charge $q \in \mathbb{Z}$ moving in the spacetime $\mathbb{R}^{24,1} \times S^1$ in presence of a $U(1)$ gauge field A_μ . We focus on a constant gauge background of the form

$$A_{25}(X) = -\frac{\theta}{2\pi R} = -i\Lambda^{-1}\partial_{25}\Lambda, \quad \Lambda(X^{25}) = \exp\left(-\frac{i\theta X^{25}}{2\pi R}\right) \quad (6.241)$$

with θ constant. The field strength is zero, so the field equations are trivially satisfied. However, the gauge parameter $\Lambda(X^{25})$ is not a global function in the spacetime $\mathbb{R}^{24,1} \times S^1$ since it does not satisfy the periodicity condition on the circle

$$\Lambda(X^{25} + 2\pi R) \neq \Lambda(X^{25}). \quad (6.242)$$

Then the background is not gauge equivalent to zero, and it has observable effects. The gauge-invariant quantity that measures the non-triviality of the background is the *Wilson line*, i.e. the holonomy of the gauge connection along the circle S^1 of charge $q \in \mathbb{Z}$,³⁹

$$W_q = \exp\left(iq \oint dx^{25} A_{25}\right) = \exp(-iq\theta) \in U(1). \quad (6.243)$$

³⁹ Or, in a better language, the *monodromy representation* $\mathbf{W}_\chi : \pi_1(\text{spacetime}) \rightarrow U(1)$ defined by the *flat* connection A in spacetime and a character $\chi : U(1) \rightarrow U(1)$. The fact that the monodromy representation captures the full gauge-invariant content of a flat connection is called the *Riemann–Hilbert correspondence*.

Let us consider a point particle of charge q with world-line action

$$\int_{\ell} d\tau \left(\frac{1}{2} \dot{X}^M \dot{X}_M - \frac{m^2}{2} - iq A_M \dot{X}^M \right) \quad (6.244)$$

The last term is just the coupling $-iq \int_{\ell} A_M dx^M$ between the world-line and the gauge connection. The path integral is a sum of integrals over paths ℓ in each homotopy class specified by the winding number⁴⁰: $Z = \sum_{w \in \mathbb{Z}} Z_w$. Switching on the flat gauge background (6.241), the contribution from the sector of winding number w picks up a Wilson line W_{wq} , i.e. $Z \rightsquigarrow \sum_{w \in \mathbb{Z}} W_{wq} Z_w$. The canonical momentum is

$$p_{25} = -\frac{\partial L}{\partial v^{25}} = v^{25} - \frac{q\theta}{2\pi R} \quad (6.245)$$

where $v^{25} \equiv i \dot{X}^{25}$. The wave-function must be periodic in X^{25} so $p_{25} = l/R$ and

$$v^{25} = \frac{2\pi l + q\theta}{2\pi R}, \quad l \in \mathbb{Z}. \quad (6.246)$$

The Hamiltonian, which annihilates the physical states, is

$$H = \frac{1}{2} (p_{\mu} p^{\mu} + v_{25}^2 + m^2) \quad (6.247)$$

and the physical mass-square of the KK modes, $-p_{\mu} p^{\mu}$, is shifted due to the dependence of v_{25} on θ

$$m_l^2 = m^2 + \left(\frac{2\pi l + q\theta}{2\pi R} \right)^2. \quad (6.248)$$

Open Strings in a Non-Trivial Flat Gauge Background

Let us now consider the bosonic open string with $U(n)$ Chan–Paton factors. A general constant A_{25} can be diagonalized by a gauge transformation

$$A_{25} = -\frac{1}{2\pi R} \text{diag}(\theta_1, \theta_2, \dots, \theta_n). \quad (6.249)$$

The background belongs to the Cartan subalgebra $\mathfrak{u}(1)^n \subset \mathfrak{u}(n)$, the Lie algebra of the maximal torus $\prod_{i=1}^n U(1)_i \subset U(n)$ of diagonal unitary matrices. Open string states belong to the adjoint $\mathfrak{u}(n)$ of $U(n)$. Hence the gauge field coupling to the Chan–Paton factor λ of a general state is

$$[A_M, \lambda], \quad (6.250)$$

⁴⁰ Indeed $\pi_1(S^1) \equiv [S^1, S^1] \simeq H^1(S^1, \mathbb{Z}) \simeq \mathbb{Z}$.

so a string in the Chan–Paton state $|ij\rangle$ has charge $+1$ for the subgroup $U(1)_i$ and charge -1 for $U(1)_j$, and is neutral for the other $U(1)_k$'s. Thus it has

$$v_{25} = \frac{2\pi l - \theta_j + \theta_i}{2\pi R}, \quad l \in \mathbb{Z}. \quad (6.251)$$

The open bosonic string mass spectrum is then

$$m^2 = \frac{(2\pi l - \theta_j + \theta_i)^2}{4\pi^2 R^2} + \frac{1}{\alpha'}(N - 1). \quad (6.252)$$

Consider the gauge bosons with $l = 0$ and $N = 1$; their mass now is

$$m^2 = \frac{(\theta_j - \theta_i)^2}{4\pi^2 R^2}. \quad (6.253)$$

In *generic* backgrounds, all the θ_i 's are distinct mod 2π , and the only massless vectors are the diagonal ones $i = j$. The unbroken gauge group in this case is $U(1)^n$, the maximal torus of $U(n)$. If k of the θ_i 's are equal, the vectors with λ in the corresponding $k \times k$ matrix block are massless, and the unbroken gauge group becomes $U(k) \times U(1)^{n-k}$. With the θ 's equal in sets of k_i the unbroken gauge symmetry is

$$U(k_1) \times \cdots \times U(k_s), \quad \sum_{i=1}^s k_i = n. \quad (6.254)$$

The low-energy interpretation of this result is obvious. From the 25-dimensional viewpoint, the gauge field A_{25} is a scalar in the adjoint representation of $U(n)$, and its v.e.v. breaks spontaneously the gauge symmetry down to the group (6.254).

Open Strings in $\mathbb{R}^{25-k,1} \times T^k$

Let us consider the open string in the spacetime $\mathbb{R}^{25-k,1} \times T^k$. From the viewpoint of the $d \equiv 26 - k$ non-compact dimensions, we have k scalars A_m transforming in the adjoint of the Chan–Paton gauge group $U(n)$. The potential of the low-energy effective action contains a term proportional to

$$\sum_{m,n} \text{tr}(F_{mn}^2) = \sum_{m,n} \text{tr}([A_m, A_n]^2), \quad (6.255)$$

which forces the gauge fields in different compactified directions to commute in a static ($\equiv$ vacuum) configuration, hence the A_m are *all simultaneously diagonalizable*.

In a compactification of the closed + open bosonic string on $T^k \equiv (S^1)^k$, we get kn moduli from the gauge field, corresponding to the n eigenvalues of the k (adjoint representation) Wilson lines

$$W_s = \exp\left(-\oint_{S_s^1} A\right), \quad s = 1, \dots, k \quad (6.256)$$

which wrap the fundamental circles of T^k , in addition to the k^2 moduli coming from the internal metric G_{mn} and antisymmetric tensor B_{mn} .

6.8 Open Bosonic String: T -Duality

We return to the open string moving in $\mathbb{R}^{24,1} \times S^1$ and consider the $R \rightarrow 0$ limit of the open string spectrum (6.252). Open strings with Neumann b.c. have no conserved quantum number similar to w ; topologically they can always be unwound from the periodic direction. Indeed, from **BOX 6.3**, we see that w is *not* conserved when the world-sheet Σ has boundary components not associated to **in/out** states. When $R \rightarrow 0$ the states with non-zero compact momentum go to infinite mass, but there is no new continuum spectrum of winding light states. The $R \rightarrow 0$ limit behaves as in field theory; the open string states propagate in 25 spacetime dimensions. This gives us a puzzle since the theories with open strings necessarily also contain closed strings which, in the $R \rightarrow 0$ limit, move in 26 spacetime dimensions, while the open strings propagate only in 25 directions.

The local 2d physics in the interior of the open and closed strings is the same, and should describe vibrations in all 26 dimensions. The open strings differ from the closed ones only because of the boundary condition at their endpoints. The only possible solution to our puzzle is that, while the string does vibrate in 26 dimensions, its endpoints are constrained to move in a 25-dimensional hyperplane in the 26d spacetime. Let us see how this happens. The T -dual string theory is most clearly described in terms of the embedding coordinate in the dual spacetime

$$X'^{25}(z, \bar{z}) = X_L^{25}(z) - X_R^{25}(\bar{z}) \quad (6.257)$$

which is the coordinate which becomes non-compact as $R \rightarrow 0$. Then

$$\partial_n X^{25} = -i \partial_t X'^{25}, \quad (6.258)$$

where n is the normal and t the tangent direction to the boundary. The Neumann boundary condition on the original coordinate $X^{25}(z, \bar{z})$ thus becomes a *Dirichlet boundary condition* on the dual coordinate $X'^{25}(z, \bar{z})$: that is, the X'^{25} coordinate of each string endpoint is *constant in time*, i.e. it is *fixed*. Thus, in terms of the T -dual coordinate X'^{25} , the string endpoints do move in the hyperplane

$$X'^{25} = \text{const.} \quad (6.259)$$

The Case of No Wilson Lines

Let us first consider compactifications *without* Wilson lines, i.e. with a trivial gauge field background. As we argued above, in the dual coordinate X'^{25} all endpoints are constrained to lie *on the same hyperplane*. To check this assertion, we compute

$$\begin{aligned} X'^{25}(\pi) - X'^{25}(0) &= \int_0^\pi d\sigma^1 \partial_1 X'^{25} = -i \int_0^\pi d\sigma^1 \partial_2 X'^{25} = \\ &= -2\pi\alpha' v^{25} = -\frac{2\pi\alpha' l}{R} = -2\pi l R', \end{aligned} \quad (6.260)$$

that is, the total variation in X'^{25} between the two ends of the string is an integral multiple of the dual period $2\pi R'$, so the two ends lie on the same hyperplane in the periodic T -dual spacetime. For two *distinct* open strings, consider the connected world-sheet Σ that results from graviton exchange between them. The argument (6.260) applies between any two boundaries of Σ , so all endpoints of *all* strings lie in the *same* hyperplane. The endpoints are still free to move in the other 24 spatial dimensions. A fancier argument: the pair

$$(dX'^{25}/2\pi R', X'^{25}|_{\partial\Sigma}) \in \Omega^1(\Sigma) \oplus \Omega^0(\Sigma) \quad (6.261)$$

represents an integral relative class in $H^1(\Sigma, \partial\Sigma; \mathbb{Z})$ [8], while a path ℓ which connects two points in $\partial\Sigma$ is a relative singular cycle. Thus $\int_\ell dX'/2\pi R'$ is an integer.

From Eq. (6.260), we see that the open string is wrapped l times on the dual periodic direction, so in the T -dual picture the open string has a conserved winding number l , but not a conserved momentum quantum number since the fixed hyperplane breaks translational invariance in the compact direction. This is the situation dual to the one in the original description.

Adding Wilson Lines

Now we switch on a flat gauge background. Due to the shift (6.251) in v_{25} , the difference (6.260) becomes

$$\Delta X'^{25} \equiv X'^{25}(\pi) - X'^{25}(0) = -(2\pi l - \theta_j + \theta_i) R'. \quad (6.262)$$

In other words, up to an additive shift, the endpoint in the Chan–Paton state i is at⁴¹

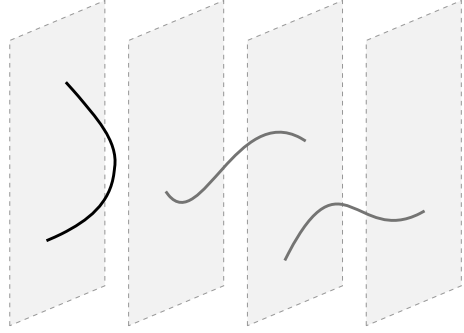
$$X'^{25} = \theta_i R' = -2\pi\alpha' A_{25,ii}. \quad (6.263)$$

Thus there are in general n distinct parallel hyperplanes at different positions in the X'^{25} -direction as in Fig. 6.2.

The mode expansion in the sector with Chan–Paton state $|i, j\rangle$ for an open string winding l times around the S^1 , is ($z = e^{\sigma_2 + i\sigma_1}$)

⁴¹ In the second equality we use Eq. (6.249).

Fig. 6.2 Open strings whose endpoints are constrained to lay on parallel hyperplanes, that is, ending on parallel D-branes in the T -dual periodic spacetime (cf. Sect. 6.9). Black (resp. dark gray) strings have the two ends on the same (resp. different) brane



$$\begin{aligned}
 X'^{25} &= \theta_i R' - \frac{i R'}{2\pi} (2\pi l - \theta_j + \theta_i) \log(z/\bar{z}) + i \left(\frac{\alpha'}{2} \right)^{1/2} \sum_{\substack{m=-\infty \\ m \neq 0}}^{\infty} \frac{\alpha_m^{25}}{m} (z^{-m} - \bar{z}^{-m}) \\
 &= \theta_i R' + \frac{\sigma^1}{\pi} \Delta X'^{25} + (2\alpha')^{1/2} \sum_{m \neq 0} \frac{\alpha_m^{25}}{m} \exp(-m\sigma^2) \sin(m\sigma^1).
 \end{aligned} \tag{6.264}$$

The mass spectrum (6.252) becomes

$$m^2 = \left(\frac{\Delta X'^{25}}{2\pi\alpha'} \right)^2 + \frac{1}{\alpha'} (N - 1), \tag{6.265}$$

where $\Delta X'^{25}$ is the minimal length (6.262) of a string in the given winding sector.

Compactification on $(S^1)^k$

Again, we may replace the original periodic coordinates X^m with the dual coordinates X'^m . The Neumann boundary condition on the X^m becomes the Dirichlet boundary condition on the X'^m , and the open string endpoints are confined to n parallel $(26 - k)$ -dimensional subspaces of $\mathbb{R}^{25-k,1} \times T^k$.

Performing T -duality along the k compact directions, the Wilson lines in the several compact directions (which are all simultaneously diagonal in a vacuum configuration) become the independent transverse coordinates in T^k of each endpoint which are constrained to lay in fixed spaces of spatial dimension $25 - k$ (or $26 - k$, counting also the time direction). Of course, we can T -dualize any number $\ell \leq k$ of compact directions, getting a mixed picture where the string endpoints are constrained to be in spatial subspaces of dimension $25 - \ell$.

6.9 D-Branes

We found that an open string on a small torus is equivalent to an open string on a large torus with its endpoints constrained to lie in certain *subspaces* of spacetime. We claim that *these subspaces are new dynamical objects of string theory*.⁴²

We present a precise proof of this claim in the next paragraph, but first we give a heuristic physical rationale of *why* the claim should hold. Consider a gravitational wave incident on a hyperplane where the open string is allowed to end. One expects the hyperplane to be pushed back by the radiation pressure of the gravitational wave causing its position and shape to change.

A New Physical Object: the Dp -Brane

Consider the spectrum of massless open strings on S^1 for the generic configuration in which the θ_i are all distinct. Ignoring the tachyon, the mass (6.265) vanishes only for $N = 1$ and $\Delta X'^{25} = 0$. That is, both endpoints should be on the same hyperplane with zero winding. We have therefore the massless states and corresponding vertices⁴³

$$\alpha_{-1}^\mu |k; ii) \quad V = i \partial_t X^\mu e^{ikX}, \quad \mu = 0, 1, \dots, 24 \quad (6.266)$$

$$\alpha_{-1}^{25} |k; ii) \quad V = i \partial_t X^{25} e^{ikX} = \partial_n X'^{25} e^{ikX}. \quad (6.267)$$

These are the same massless states as in the original open string theory: T -duality gives simply a different picture of the *same* physics. We saw in Sect. 6.7 that the only massless states of the original theory with *generic* Wilson lines are the $U(1)^n$ vectors in the Cartan algebra. In Eqs. (6.266), (6.267) we have divided these states in two groups: the states (6.266) have a polarization *tangent* to the fixed hyperplane $X^{25} = \text{const}$, while the states (6.267) have polarization *perpendicular* to it.

We stress that the momentum k^μ is always *tangent to the hyperplane*, since there is no zero-mode linear in σ^2 in the mode expansion (6.264) of X'^{25} , that is, $k_{25} \equiv 0$: this means that the corresponding physical states (6.266), (6.267) *propagate only along the i th hyperplane of equation*

$$X^{25} = \theta_i R' \equiv \text{const}. \quad (6.268)$$

In other words, *the states (6.266), (6.267) describe degrees of freedom which live on the hyperplane, not in the bulk of spacetime*.

The states with polarization parallel to the hyperplane yield a 25d massless gauge vector living on the i th hyperplane. The state (6.267) with a polarization perpendicular to the hyperplane is a 25d scalar living on the i th hyperplane which has a simple interpretation in the T -dual theory; *it is a collective coordinate for the shape of the hyperplane*. This is natural in view of the fact that a constant gauge field $A_{25,ii}$ along the S^1 in the original theory corresponds to a shift of the T -dual hyperplane; cf.

⁴² General references for this section and the following ones are [57, 58].

⁴³ The polarizations of these vertices are transverse with respect to k^μ .

Eq. (6.263). A x^μ -dependent background $A_{25,ii}(x^\mu)$ then corresponds in the T -dual picture to strings ending on the curved hypersurface of equation⁴⁴

$$x'^{25} = -2\pi\alpha' A_{25,ii}(x^\mu), \quad (6.269)$$

and the quanta of the field $A_{25,ii}(x^\mu)$ correspond in the T -dual picture to small oscillations of the i th endpoint hypersurface.

In Chap. 1, we started with a string moving in flat space, and soon discovered that its spectrum contains a massless metric field $G_{\mu\nu}$ so that the spacetime geometry is dynamical. Here we first construct a flat hypersurface in spacetime, and then discover a massless field in the physical spectrum which allows the hypersurface to move around and change its shape. We conclude that the hyperplane is a *dynamical* object of the theory, called a *Dirichlet membrane*, or simply *D-brane*. The p -dimensional brane, obtained by T -dualizing $25 - p$ spatial directions, is called a **Dp-brane**. The world-story of a Dp-brane then tracks a $(p + 1)$ -dimensional submanifold $V \subset M$ in spacetime called its *world-volume* (*world-sheet* for $p = 1$). The spacetime metric $G_{\mu\nu}$ induces a metric on V of signature $(p, 1)$.

Conclusions The bosonic string theory contains, besides the fundamental string itself, other dynamical extended objects, of dimension p with $0 \leq p \leq 25$, the *Dp-branes*, which vibrate and move in the 26-dimensional spacetime. D-branes are characterized by the property that a fundamental string may end on them, that is, they are allowed boundary conditions for the world-sheet QFT on the string.

The scalars X^μ propagating on the world-sheet of a string which ends on a D-brane satisfy the Dirichlet b.c. $X^\mu = \text{const.}$ for μ a direction normal to the brane and the Neumann b.c. $\partial_\eta X^\mu = 0$ for μ tangent to the brane. The massless degrees of freedom living on the world-volume of an *isolated* Dp-brane are:

- (a) a $U(1)$ gauge vector A_μ ($\mu = 0, 1, \dots, p$)
- (b) scalars X'^m ($m = p + 1, \dots, 25$) describing transverse motions of the brane.

Effect of T -Duality on Dp-Branes Since T -duality interchanges Dirichlet and Neumann boundary conditions, a T -duality in a direction tangent to a Dp-brane reduces it to a $D(p - 1)$ -brane, while a T -duality in an orthogonal direction turns it into a $D(p + 1)$ -brane. More general cases of T -duality along a “diagonal” direction will be discussed in the next subsection.

Crucial Warning Above and below we ignore the tachyon living on the world-volume of the D-brane. This tachyon signals that the configuration is unstable, that is, that the D-brane is *not a stable object* and it may dissolve. This is physically expected since the bosonic string D-brane does not carry any conserved charge whose conservation prevents its dissolution. We stress again that in this textbook the bosonic string is meant to be a mere *kindergarten laboratory*, used for the didactical purpose of introducing techniques and tools for the superstring in their simplest possible context.

⁴⁴ The statement is exact in the limit of an adiabatic dependence of $A_{25,ii}$ on the coordinates x^μ along the hypersurface.

In the superstring case, the D -branes will have no tachyon, will be absolutely stable, and will carry non-trivial conserved charges.

D-Branes and Non-Abelian Gauge Fields

We return to Eq. (6.263). We consider string theory in the space $\mathbb{R}^{p,1} \times T^{25-p}$ with all radii equal R , perform the T -duality in all the $25 - p$ compact directions, and take the dual decompactification limit $R' \rightarrow \infty$ (i.e. $R \rightarrow 0$) to return to flat 26 dimensions while scaling the nk Wilson line moduli so that the branes transverse positions

$$X_i'^m \stackrel{\text{def}}{=} \theta_i^m R' = \text{fixed}, \quad m = p + 1, \dots, 25, \quad i = 1, 2, \dots, n, \quad (6.270)$$

are kept fixed. We get a configuration (a “*stack*”) of parallel D -branes in $\mathbb{R}^{25,1}$, all of the same dimension p , stretched in the directions X^a ($a = 1, \dots, p$) and separated in the transverse directions X^m . Note that the Chan–Paton index i now labels the different Dp -branes in the stack.

We focus on the regime where the transverse separations between the branes are small in string units

$$|X_i'^m - X_j'^m| \ll 2\pi\sqrt{\alpha'}. \quad (6.271)$$

In this regime (neglecting the tachyons), the parametrically light states propagating on the branes (i.e. with momenta constrained to be parallel to the brane stack) are

$$\alpha_{-1}^\mu |k; ij\rangle \quad \mu = 0, 1, \dots, p; \quad \alpha_{-1}^m |k; ij\rangle \quad m = p + 1, \dots, 25 \quad (6.272)$$

with $i, j = 1, \dots, n$ yielding n^2 vectors and n^2 scalars of mass

$$m_{ij}^2 = \sum_m \left(\frac{\Delta X_{ij}^m}{2\pi\alpha'} \right)^2. \quad (6.273)$$

These formulae have a simple physical interpretation: on the world-volume of the stack of parallel D -branes there lives a Yang–Mills theory with gauge group $U(n)$ coupled to $25 - p$ scalar fields X_{ij}^m in the adjoint representation. A non-zero v.e.v. of the scalar fields break the gauge group *à la* Higgs. By the argument around Eq. (6.255) in a static configuration, the v.e.v. $\langle X_{ij}^m \rangle$ are all simultaneously diagonalizable, and their eigenvalues $X_i'^m$ yield the position of the i th brane in the transverse space.

This set-up yields an interesting *geometric* re-interpretation of the $U(n)$ gauge symmetry breaking (the Higgs phenomenon) in the T -dual picture. When no D -branes are on top of each other (i.e. $X_i'^m \neq X_j'^m$ for $i \neq j$), there is just one massless vector on each brane, producing the unbroken gauge group $U(1)^n$. If r D -branes coincide, there are new massless states because strings stretched between these branes have vanishing length $\Delta X'^{25} = 0$ when the endpoints i and j belong to the same set of r branes on top of each other. *This is the same physics as found in the dual*

Wilson line picture. In the brane description, we have also r^2 massless scalars from the components of A^M normal to the D-branes. We shall discuss them below.

In particular, the Abelian vector associated to central subgroup $U(1) \subset U(n)$ does not couple to the adjoint matter $X'_{ij}{}^m$ and hence is infra-red free. Likewise the neutral scalar $\text{tr}(X'^m)$ is IR free; geometrically it corresponds to the overall translation of the center of mass of the branes' stack, which has obviously a trivial dynamics by target-space Poincaré symmetry.

Reinterpretation of Open String Theory In the language of this section, the original (oriented) open string theory with CP gauge group $U(n)$ corresponds to the configuration containing n D25-branes which with number as $i = 1, \dots, n$. In a D25-brane all fields X^μ have Neumann b.c. This is exactly the boundary condition we used in the original definition of the open string: a D25-brane fills spacetime, so the string end-points may be anywhere in spacetime. The Chan–Paton index i of a string endpoint is the number i of the brane on which it lays. The D25-brane viewpoint thus yields a deeper understanding of the Chan–Paton degree of freedom. In particular, as for all Dp -branes, we should expect the D25 to have a non-zero tension, $\tau_{25} \neq 0$, which yields a physical interpretation to the disk dilaton tadpole discovered in Sect. 4.5.1. See Eq. (6.309) below. On the other hand, we expect the D25 to be unstable, as all Dp -branes. This instability explains the presence of a tachyon in the open bosonic string [59].

Emergence of Non-Commutative Geometry If we have a stack of n parallel Dp -branes, their transverse coordinates X_{ij}^m are $n \times n$ matrices. In a static configuration, these matrices commute, may be simultaneously diagonalized, and the eigenvalue x_i^m of X_{ij}^m is interpreted as the position of the i th brane in the m th transverse direction. However, for a general non-static configuration, the branes' coordinates do *not* commute. This is a first hint of the fact that string theory sees spacetime not as a classical (i.e. *commuting*) manifold but as a geometry which is non-commutative at the string length scale $\sqrt{\alpha'}$. We shall find other hints of this deep fact. Again we recognize that string theory sees the short-distance structure of spacetime in a way which is radically different from conventional QFT; this should be expected since string theory is a UV-finite theory of quantum gravity and must detect the quantum nature of spacetime at short distances. For non-commutative geometry and string theory, see [60].

6.9.1 D-Brane Action (Bosonic String)

We saw that the massless fields living on the $(p + 1)$ -dimensional world-volume V of an isolated Dp -brane are a $U(1)$ gauge vector field and $25 - p$ world-volume scalars X^m describing its normal fluctuations. We now consider the low-energy effective action for these light modes. As before, we take the dual radius $R' \rightarrow \infty$, and focus on the D-brane in 26 non-compact flat dimensions. The modes living on the brane interact with the bulk string fields propagating in 26 dimensions whose action is

described in Eq. (6.148). We write ξ^a ($a = 0, 1, \dots, p$) for the coordinates on the brane. As always in this chapter, we ignore all tachyons.⁴⁵ The remaining light fields living on the brane are its embedding coordinates $X^\mu(\xi)$ and the gauge field $A_a(\xi)$.

Action of an Isolated Dp -Brane

We claim that the world-volume effective action for an *isolated* brane moving in a 26d background of the bulk massless field $G_{\mu\nu}$, $B_{\mu\nu}$, Φ is simply

$$S_p = -T_p \int_V d^{p+1}\xi \, e^{-\Phi} \left[-\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab}) \right]^{1/2}, \quad (6.274)$$

where T_p is a constant (the Dp -brane *tension*) to be computed later, and

$$G_{ab} \equiv \frac{\partial X^\mu}{\partial \xi^a} \frac{\partial X^\nu}{\partial \xi^b} G_{\mu\nu}, \quad B_{ab} \equiv \frac{\partial X^\mu}{\partial \xi^a} \frac{\partial X^\nu}{\partial \xi^b} B_{\mu\nu} \quad (6.275)$$

are the induced ($\equiv$ pulled-back) metric and B -field on the brane world-volume V . F_{ab} is the field strength of the Abelian gauge field A_a living on the brane.

All features of the action (6.274) can be justified by simple physical arguments. Consider first a configuration in which the only non-trivial bulk field is the metric while the $U(1)$ gauge field vanishes. The lowest-derivative coordinate-invariant action is the integral of $(-\det G_{ab})^{1/2}$ which is just the volume of the world-volume $V \subset M$ as a submanifold of spacetime, i.e. the generalized Nambu–Goto action. The dilaton dependence $e^{-\Phi} \propto g_c^{-1}$ arises because (6.274) is an open string tree-level action, so computed by disk amplitudes, and $\chi(D) = 1$.

The dependence on F_{ab} follows from T -duality. Consider a D -brane which is extended in the directions X^1 and X^2 with a constant gauge field F_{12} on it. We use the gauge

$$A_2 = X^1 F_{12} \quad (6.276)$$

and take the T -dual along the 2-direction. The Neumann boundary condition in this direction becomes Dirichlet, so the D -brane loses one dimension. The T -duality relation (6.263) between the potential and the dual coordinate yields

$$X^{2'} = -2\pi\alpha' F_{12} X^1, \quad (6.277)$$

that is, the T -dual D -brane is tilted in the $(1-2)$ -plane by an angle θ related to the magnetic field F_{12} on the world-volume of the original brane by the equation

$$\theta = -\arctan(2\pi\alpha' F_{12}). \quad (6.278)$$

The slope $\tan \theta$ yields a geometric factor in the NG action

⁴⁵ Hence our treatment is at best purely formal. The conclusions will however apply to the superstring D -branes and will be very robust in that setting.

$$\int dX^1 [1 + (\partial_1 X'^2)^2]^{1/2} = \int dX^1 [1 + (2\pi\alpha' F_{12})^2]^{1/2}. \quad (6.279)$$

In presence of a general constant field strength F_{ab} on V , we go to a frame where F_{ab} is skew-diagonal, and we reduce to the previous 2-dimensional case, getting the action (6.274) (with $B = 0$). The determinant form of the action in Eq. (6.274) for the gauge field A_a is known as the *Born-Infeld-(Dirac) action* (BI) [61, 62]. Contrary to the vast majority of Lagrangians which are higher order in the field strength F_{ab} , the BI is fully consistent with causality and unitarity [63].

The dependence of the action from the B -field is fixed by the following version of the Higgs mechanism. The closed string field $B_{\mu\nu}$ and the open string A_μ appear in the fundamental *string* world-sheet action through the geometric couplings

$$\frac{i}{2\pi\alpha'} \int_\Sigma B + i \int_{\partial\Sigma} A. \quad (6.280)$$

Associated to each of these fields there is a spacetime gauge symmetry which should be preserved if the physical theory should be consist. The usual gauge transformation

$$\delta A = d\lambda \quad (6.281)$$

is an invariance of the action (6.280) since

$$\delta \int_{\partial\Sigma} A = \int_{\partial\Sigma} d\lambda = \int_{\partial^2\Sigma} \lambda = 0 \quad (6.282)$$

because the boundary of a boundary is empty. The B -field variation

$$\delta B = d\zeta \quad (6.283)$$

instead changes the world-sheet bulk action by a boundary term

$$\delta \int_\Sigma B = \int_{\partial\Sigma} \zeta. \quad (6.284)$$

The variation (6.284) can be canceled only if the open string field A also transforms under the B -field gauge transformation as

$$\delta A = -\frac{1}{2\pi\alpha'} \zeta. \quad (6.285)$$

Only the combination

$$2\pi\alpha' \mathcal{F}_{\mu\nu} \stackrel{\text{def}}{=} B_{\mu\nu} + 2\pi\alpha' F_{\mu\nu} \quad (6.286)$$

is fully invariant under both gauge symmetries, and this is the combination which should enter in the action. The form (6.274) is thus completely fixed by these general

physical considerations up to the overall constant T_p . One may check our findings by computing explicitly the relevant string amplitudes.

Exercise 6.9 Check our claim (6.274) by computing the appropriate amplitudes of the various massless BRST-invariant vertices on a disk with the appropriate N/D b.c.

Stacks of Parallel Dp -Branes

For n well-separated D -branes the action is n copies of the single brane action. However we have seen that when the D -branes are on top of each other, there are n^2 instead of n massless vectors and scalars propagating along the brane, and we would like to write the effective action which governs them. The fields A_μ and A_m are now $n \times n$ matrices. For the vector field, the meaning of this fact is obvious—the A_μ become $U(n)$ non-Abelian gauge fields. For the collective coordinate X^μ , however, the interpretation is less easy. The collective coordinates for the embedding of the n D -branes in spacetime are promoted to $n \times n$ matrices. As already advertised, this leads to *non-commutative geometry* in the D -brane sector of string theory.

Now there is also a potential term for the collective coordinates X^m which can be deduced by T -duality from the YM action. For constant gauge fields, we have

$$F_{mn} = [A_m, A_n] \quad (6.287)$$

which in the T -dual picture becomes (cf. Eq. (6.263))

$$F_{mn} \rightsquigarrow \frac{1}{(2\pi\alpha')^2} [X_m, X_n]. \quad (6.288)$$

From the kinetic term $\text{tr } F_{mn}^2$ in the YM action, we get the potential term on the brane

$$V(X^m) = \frac{1}{(4\pi g_{\text{YM}}\alpha')^2} \text{tr}([X_m, X_n][X^m, X^n]) + \text{higher order in } [X^m, X^n]. \quad (6.289)$$

This potential has the property that at the point $X^m = 0$ has vanishing second derivative, so all $(25 - p)n^2$ scalars are massless. However the dimension of the space of flat directions in field space is smaller; the potential vanishes if and only if $[X^m, X^n] = 0$ for all m, n . Then we can go to a gauge where all X^m are *simultaneously* diagonal. Thus—modulo gauge equivalence—there are just n flat directions for each of the $(25 - p)$ directions orthogonal to the brane given by the eigenvalues of the matrices X^m . These n eigenvalues are then interpreted as transverse collective coordinates for the n Dp -branes. The potential (6.289) interpolates between the physics of the well-separated and coincident branes. When the X^m commute, the effective action should reduce to the sum of n separate D -branes, hence at low-energy

$$S_p = -T_p \int d^{p+1} \xi \left\{ e^{-\Phi} \sqrt{-\det(G_{ab} + 2\pi\alpha' \mathcal{F}_{ab})} + O([X^m, X^n]^2) \right\}. \quad (6.290)$$

The determinant is on the world-indices a, b and the trace on the Chan–Paton indices.

D-Brane Tension

Before computing the Dp -brane tension, we prove an a priori recursion relation in p for T_p . We wrap the Dp -brane on the p -torus T^p of a $\mathbb{R}^{25-p,1} \times T^p$ geometry by identifying periodically its world-volume spatial directions X^m . From the viewpoint of the non-compact $(26 - p)$ -dimension, this configuration looks like a particle of mass equal to its effective tension $T_p e^{-\Phi}$ times the volume of the torus

$$\text{mass} = T_p e^{-\Phi} \prod_{i=1}^p (2\pi R_i). \quad (6.291)$$

Now take the T -dual of the last periodic dimension X^p . This does not change the mass, since it is just a different description of the same physical state. In terms of the dilaton of the T -dual picture

$$e^{-\Phi'} = (\alpha')^{-1/2} e^{-\Phi} R_p, \quad (\text{cf. Eq. (6.124)}) \quad (6.292)$$

the mass (6.291) becomes

$$\text{mass} = 2\pi \alpha'^{1/2} T_p e^{-\Phi'} \prod_{i=1}^{p-1} (2\pi R_i). \quad (6.293)$$

In the T -dual theory, this state is seen as a $D(p - 1)$ -brane wrapped on a $(p - 1)$ -torus so its mass is

$$\text{mass} = T_{p-1} e^{-\Phi'} \prod_{i=1}^{p-1} (2\pi R_i). \quad (6.294)$$

Equating the two expressions, we get

$$T_p = \frac{T_{p-1}}{2\pi (\alpha')^{1/2}}, \quad (6.295)$$

a recursion which determines all T_p 's in terms of one of them.

Direct Computation of the Tension from the Cylinder Amplitude

To compute directly the tension, we consider two parallel Dp -branes at positions $X_1^m = 0$ and $X_2^m = y^m$. These two objects can feel each other presence by exchanging closed strings through the process in Fig. 6.3: this is the dual-channel re-interpretation of the one-loop amplitude for open strings whose endpoints are constrained to lay on the two branes. The string world-sheet Σ for this process is a cylinder, with no operator inserted, whose two boundaries are mapped, respectively, to the two parallel branes. In other words: the world-sheet scalars X^m in the directions perpendicular to the branes satisfy the Dirichlet boundary condition which fixes their values on the boundaries to be X_1^m and X_2^m , respectively.



Fig. 6.3 The open string amplitude which computes the D-brane tension. From the open sector perspective, it is a one-loop trace over the Hilbert space of open strings whose endpoints are constrained to lay on the two branes. From the crossed-channel viewpoint, it is a tree-level exchange of closed string states between the two branes

Looking at the cylindrical amplitude in Fig. 6.3 from the open string channel, we see that it is given by a trace in the open string Hilbert space as in Chaps. 4 and 5. The residues at the graviton and dilaton poles in the crossed-channel then yield the coupling T_p we are looking for [57, 64, 65]. The expression for the cylindrical amplitude is again given by Eq. (4.181),

$$2 \int_0^\infty \frac{dt}{t} \text{Tr}'_{12} [\exp(-2\pi t (L_0 - 1))], \quad (6.296)$$

where the trace is on the physical ($\equiv$ transverse) Hilbert space $\mathbf{H}_{12}$ of open strings in the sector with the appropriate boundary conditions. The scalars X^m transverse to the branes satisfy the fixed Dirichlet b.c.

$$X^m(0) = 0 \quad X^m(\pi) = y^m, \quad (6.297)$$

while the tangent scalars X^μ satisfy the usual Neumann b.c. The overall factor 2 in Eq. (6.296) arises from the sum over the two orientations of the open string. We get the same amplitude as in Eq. (4.181) except for the following obvious modifications:

- the number of momentum integrations (i.e. integrations over the zero-modes of non-compact scalars) is reduced from 26 to $p + 1$ and dually V_{26} becomes V_{p+1} ;
- the Virasoro weight h_i ($\equiv$ the eigenvalue of L_0) acquires an additional term

$$\Delta h_i = \frac{y^2}{4\pi^2\alpha'}, \quad \left(\text{where } y^2 \equiv y^m y^m \right) \quad (6.298)$$

from the tension of the stretched open string (cf. Eq. (6.265));

- the Chan–Paton weight n^2 is no longer present since we are not summing over all possible boundary states; the Chan–Paton labels of the two boundaries now are

fixed. Instead we have a factor 2 since the string can attach to the branes with both orientations.

Putting all ingredients together, we get

$$\begin{aligned} i V_{p+1} \int_0^\infty \frac{dt}{t} (8\pi^2 \alpha' t)^{-(p+1)/2} e^{-ty^2/2\pi\alpha'} \eta(it)^{-24} = \\ = \frac{i V_{p+1}}{(8\pi^2 \alpha')^{(p+1)/2}} \int_0^\infty dt t^{(21-p)/2} e^{-ty^2/2\pi\alpha'} \left(e^{2\pi/t} + 24 + \dots \right), \end{aligned} \quad (6.299)$$

BOX 6.12 - The scalar Green's function in d -dimensions

By definition

$$G_d(y) = \int \frac{d^d p}{(2\pi)^d} \frac{e^{ip \cdot y}}{p^2} = \int \frac{d^d p}{(2\pi)^d} \int_0^\infty dt e^{-p^2 t + i p \cdot y}.$$

Performing the Gaussian integral in p

$$G_d(y) = \frac{1}{(2\pi)^d} \int_0^\infty dt \left(\frac{\pi}{t} \right)^{d/2} e^{-y^2/4t} = \frac{y^{2-d}}{4\pi^{d/2}} \int_0^\infty ds s^{d/2} e^{-s}$$

where we set $s = y^2/4t$. Then

$$G_d(y) = \frac{1}{4\pi^{d/2}} \Gamma\left(\frac{d}{2} - 1\right) |y|^{2-d}.$$

where we used the expansion of $\eta(\tau)$ in Sect. 4.7. The first term in the large parenthesis arises from the tachyon exchange and is not interesting for our present purposes; the integral over the second term yields

$$\begin{aligned} i V_{p+1} \frac{24}{2^{12}} (4\pi^2 \alpha')^{11-p} \pi^{(p-23)/2} \Gamma\left(\frac{23-p}{2}\right) |y|^{p-23} = \\ = i V_{p+1} \frac{24\pi}{2^{10}} (4\pi^2 \alpha')^{11-p} G_{25-p}(y) \end{aligned} \quad (6.300)$$

where $G_d(y)$ is the massless scalar Green's function in d dimensions, see BOX 6.12.

Now we compare this stringy computation with a field-theoretic evaluation [65] of the same amplitude. The B -field does not couple to the D-branes, so the relevant terms in the effective spacetime action (1.168), written in the Einstein frame, are

$$\frac{1}{2\kappa^2} \int d^{26}x \sqrt{-G} \left(R - \frac{1}{6} \nabla_\mu \Phi \nabla^\mu \Phi \right) \quad (6.301)$$

The dilaton has been shifted so that its v.e.v. vanishes. In terms of the same fields, the relevant terms from the D-brane action (6.274) are

$$- \tau_p \int d^{p+1} \xi \exp\left(\frac{p-11}{12} \Phi\right) \sqrt{-\det G_{ab}} \quad (6.302)$$

where we set $\tau_p = T_p e^{-\Phi_0}$; this is the physical tension of the Dp -brane when the dilaton background value is Φ_0 .

The field theory counterpart to Fig. 6.3 is the exchange graph of a single graviton or dilaton between the D-branes. To obtain the propagator, we expand the spacetime action to second order in $h_{\mu\nu} \equiv G_{\mu\nu} - \eta_{\mu\nu}$ and in Φ . In addition, we need to impose a gauge condition for the gravity field; the most convenient one is

$$F_\rho \equiv \eta^{\mu\nu} \partial_\mu h_{\nu\rho} - \frac{1}{2} \partial_\rho (\eta^{\mu\nu} h_{\mu\nu}) = 0. \quad (6.303)$$

Adding to the action the gauge-fixing term $-\eta^{\mu\nu} F_\mu F_\nu / 4\kappa^2$, to the second order in the fluctuations the action reads

$$- \frac{1}{8\kappa^2} \int d^{26}x \left(\partial_\mu h_{\nu\lambda} \partial^\mu h^{\nu\lambda} - \frac{1}{2} \partial_\mu (h^\nu_\nu) \partial^\mu h^\lambda_\lambda + \frac{2}{3} \partial_\mu \Phi \partial^\mu \Phi + \dots \right), \quad (6.304)$$

where the indices are raised with the flat metric $\eta^{\mu\nu}$. Inverting the propagators, one obtains the momentum space propagators (we write them for general dimension D)

$$\langle \Phi \Phi \rangle = - \frac{(D-2)i\kappa^2}{4k^2} \quad (6.305)$$

$$\langle h_{\mu\nu} h_{\sigma\rho} \rangle = - \frac{2i\kappa^2}{k^2} \left(\eta_{\mu\sigma} \eta_{\nu\rho} + \eta_{\mu\rho} \eta_{\nu\sigma} - \frac{2}{(D-2)} \eta_{\mu\nu} \eta_{\sigma\rho} \right). \quad (6.306)$$

The D-brane action expanded around a flat background is

$$- \tau_p \int d^{p+1} \xi \left(\frac{p-11}{12} \Phi - \frac{1}{2} h_{aa} \right) \quad (6.307)$$

Note that $h_{\mu\nu}$ is traced only over the directions tangent to the D-brane; we have taken ξ to be flat coordinates with metric δ_{ab} .

We can now read off the Feynman graph amplitude [57, 65]

$$\frac{i\kappa^2 \tau_p^2}{k_\perp^2} V_{p+1} \left\{ 6 \left[\frac{p-11}{12} \right]^2 + \frac{1}{2} \left[2(p+1) - \frac{1}{12} (p+1)^2 \right] \right\} = \frac{6i\kappa^2 \tau_p^2}{k_\perp^2} V_{p+1} \quad (6.308)$$

Comparison with the stringy result (6.300) then yields [64]

$$T_p^2 = \frac{\pi}{256 \kappa^2} (4\pi^2 \alpha')^{11-p} \quad (6.309)$$

As a check, we see that the recursion relation (6.295) is satisfied.⁴⁶

6.10 *T*-Duality of Unoriented Strings: Orientifolds

Finally, we take the $R \rightarrow 0$ limit of *unoriented* strings.⁴⁷

Closed Unoriented Strings

To get the unoriented theory, we impose the projection $\Omega = +1$ and see this procedure as *gauging* the orientation-reversing symmetry Ω . The *T*-dual set-up is best described using the dual coordinates

$$X'^m(z, \bar{z}) = X_L^m(z) - X_R^m(\bar{z}) \quad (6.310)$$

instead of the original ones

$$X^m(z, \bar{z}) = X_L^m(z) + X_R^m(\bar{z}). \quad (6.311)$$

Here and below m labels the *T*-dualized directions, μ the non-dualized ones.

In the original description, we are gauging the world-sheet parity Ω which acts as

$$\Omega: X_L^M(z) \leftrightarrow X_R^M(\bar{z}). \quad (6.312)$$

In terms of the *T*-dual coordinates, this is

$$X'^m(z, \bar{z}) \leftrightarrow -X'^m(\bar{z}, z), \quad X'^\mu(z, \bar{z}) \leftrightarrow X'^\mu(\bar{z}, z). \quad (6.313)$$

Thus, in the dual coordinates, Ω becomes the product of world-sheet parity with a spacetime reflection. We know that gauging world-sheet parity alone produces the unoriented theory, while gauging the reflection alone produces a $\mathbb{Z}_2$ orbifold. The combined gauging produces a new object known as the *orientifold*.

Orientifolds

To a certain extent, orientifolds are similar to orbifolds, but there are also crucial differences. We separate the string wave-functions in the dependence on the center of mass x^m and its internal d.o.f., and take the internal part of the wave-function to be an eigenstate of Ω . The overall projection on $\Omega = +1$ states then sets the string wave-function at $-x^m$ and at x^m to be equal up to sign. For instance, the components of the metric and antisymmetric tensors satisfy

⁴⁶ Note that the equality of the mass computed in the two ways yields another proof for the transformation of the dilaton under *T*-duality.

⁴⁷ References for this section are [66–71] and the reviews [57, 72, 73].

$$\begin{aligned}
G_{\mu\nu}(x') &= G_{\mu\nu}(x) & B_{\mu\nu}(x') &= -B_{\mu\nu}(x) \\
G_{\mu n}(x') &= -G_{\mu n}(x) & B_{\mu n}(x') &= B_{\mu n}(x) \\
G_{mn}(x') &= G_{mn}(x) & B_{mn}(x') &= -B_{mn}(x)
\end{aligned} \tag{6.314}$$

where

$$(x^\mu, x^m)' = (x^\mu, -x^m). \tag{6.315}$$

In Eq. (6.314) there is a minus sign for each m, n index and an additional minus for the 2d parity-odd B -field; in the orbifold case the last sign is absent.

The T -dual spacetime is the torus T^k modded out ($\equiv$ identified) under a $\mathbb{Z}_2$ reflection of the k compact directions, just as in the orbifold construction. E.g. for $k = 1$ the dual spacetime is $\mathbb{R}^{24,1} \times I$ with I the line segment

$$0 \leq x^{25} \leq \pi R' \tag{6.316}$$

with *orientifold fixed planes at the two ends*. We stress that away from the orientifold fixed planes that the local physics is the one of *oriented* string theory. Unlike the original unoriented theory where the Ω -projection removes locally half the states, here it relates the string wave-function at one point x with the string wave-function at its image point x' ; cf. Eq. (6.315) without imposing local restrictions. An important difference between the orbifold and orientifold constructions is that the latter has no analog of twisted states.⁴⁸

It should be stressed that *the orientifold planes cannot be dynamical objects*. Unlike the D -branes, there are no string modes which represent fluctuations of the orientifold shape and position. The heuristic argument saying that a gravitational wave incident on a D -brane forces it to oscillate, does not apply to orientifold planes; the identifications (6.314) become boundary conditions at the fixed plane which force the incident and reflected gravitational waves to cancel each other. On the contrary, in the case of the D -brane the reflected wave gets small at weak string coupling.

Open Unoriented Strings

For simplicity, we focus on the case of one compact dimension. Again there is one orientifold fixed plane at $X'^{25} = 0$ and another one at $X'^{25} = \pi R'$. Introducing (say) $SO(n)$ Chan–Paton factors, a general Wilson line can be set in the diagonal form; for n even the eigenvalues are paired

$$W = \text{diag}(e^{i\theta_1}, e^{-i\theta_1}, e^{i\theta_2}, e^{-i\theta_2}, \dots, e^{i\theta_{n/2}}, e^{-i\theta_{n/2}}) \tag{6.317}$$

Thus in the dual picture, there are $n/2$ D -branes on the line segment

$$0 \leq X'^{25} \leq \pi R' \tag{6.318}$$

⁴⁸ This is consistent with the fact that Klein bottle has no analogue of the modular transformation S which in the orbifold case required the existence of twisted states.

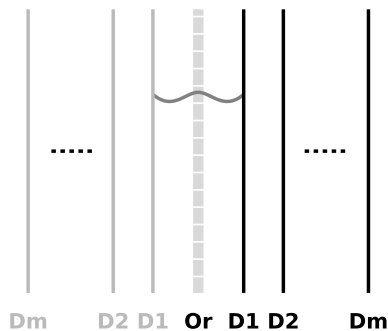


Fig. 6.4 m D-branes near and parallel to the orientifold plane **Or**. Branes/orientifold are extended in the vertical direction and in $(p - 1)$ directions orthogonal to the page. Branes (resp. *image branes*) are drawn black (resp. *light gray*). The dark gray curve is a fundamental string which stretches between a brane and its mirror image on the other side of **Or**

and $n/2$ at their image points under the orientifold identification. Strings can stretch between D-branes and their images as in Fig. 6.4

The generic unbroken gauge group is $U(1)^{n/2}$. As in the oriented case, if r D-branes are coincident the unbroken gauge group is $U(r) \times U(1)^{n/2-r}$. However, now we have new configurations: if r D-branes lie on one of the fixed planes, then strings stretching between one of these branes and one of the image branes also become massless, and the unbroken gauge group contains a $SO(2r)$ factor. The maximal unbroken gauge group, $SO(n)$, is recovered if all the branes are stacked on top the orientifold plane. Note that this maximally symmetric case is asymmetric with respect to the two fixed planes in the geometry.

When n is odd the last eigenvalue is ± 1 so that in the T -dual picture a brane should lie on one of the two fixed planes; since it has no image, this is actually a *half* D-brane as measured (say) by its tension. If we consider for $n = 2$ the Wilson line $\text{diag}(1, -1)$ —which is an element of $O(2)$ but not of its connected subgroup $SO(2)$ —rather than one D-brane and its image, we have two half D-branes one on each fixed plane.

The orientifold plane couples to the dilaton and the metric. The amplitude is as in the previous section with the Klein bottle and the Möbius strip replacing the cylinder in Fig. 6.3. We already did the relevant computation in Sect. 4.9; there we found that for the original non-oriented open strings the total dilaton coupling cancels iff the gauge group is $SO(2^{13})$. In the T -dual picture, the total dilaton coupling of the orientifold planes cancels that of 2^{12} D-branes (we don't count images!). More generally, if we T -dualize $k = 25 - p$ dimensions, we get 2^k fixed planes; thus the effective action for a single fixed plane is

$$2^{12-k} T_p \int d^{p+1} \xi e^{-\Phi} \sqrt{-\det G_{ab}} \quad (6.319)$$

where the integral is over the fixed plane.

Appendix 1: Non-Abelian Orbifolds

We consider a string moving on a target-space M , i.e. the world-sheet theory is a 2d σ -model with target space M (possibly with other couplings such as the Wess-Zumino term) which is conformal invariant. As before, we identify the field configurations we integrate over in the path integral with the maps $X: M \rightarrow M$. Let G be a discrete subgroup of the symmetries of the 2d QFT (so, in particular, $G \subset \text{Iso}(M)$, the isometry group of M). We wish to discuss the string (or more generally the CFT) with target space the orbifold $X \equiv M/G$. We are interested in the case that Γ acts non-freely, otherwise X is a smooth manifold. As in Sect. 6.6 we can consider the *twisted sectors* $\mathbf{H}_g$ of the Hilbert space corresponding to the string configurations which satisfy the twisted boundary conditions

$$X(2\pi) = gX(0), \quad g \in G. \quad (6.320)$$

The case that G is Abelian was already discussed in Sect. 6.6. When G is non-Abelian the situation is a bit more tricky [45, 46] because some group element takes one twisted sector to another. Indeed acting with $h \in G$ in the sector with b.c. (6.320)

$$X(0) \rightsquigarrow hX(0) \quad \text{and} \quad X(2\pi) \rightsquigarrow hX(2\pi) = hgX(0) = hgh^{-1}(hX(0)) \quad (6.321)$$

so that $\mathbf{H}_g \rightsquigarrow \mathbf{H}_{hgh^{-1}} \neq \mathbf{H}_g$ if g and h do not commute. That is: the group action permutes the sectors $\mathbf{H}_g$ with g in the same G -conjugacy class. Thus to form G -invariant states we have to take a state $|\psi\rangle_g \in \mathbf{H}_g$ in the g -twisted sector, project it onto the invariant subspace of the centralizer $C(g) \subset G$ of g , and then take the sum of corresponding states from sectors in the same conjugacy class C_g , each of which is projected onto its centralizer-invariant subspace

$$|\psi, \text{inv.}\rangle_g = \frac{1}{|C_g|} \sum_{f \in C_g} \frac{1}{|C(f)|} \sum_{h \in C(f)} h|\psi\rangle_f \quad (6.322)$$

so we have one Hilbert space sector per conjugacy class in G . The one-loop partition function corresponding to the projection (6.322) is

$$Z(\tau)_{M/G} = \frac{1}{|G|} \sum_{\substack{g, h \in G \\ gh=hg}} Z_{h,g}(\tau) \quad (6.323)$$

where $Z_{h,g}(\tau)$ is the path integral over the fields on the torus of modulus τ with twisted b.c.

$$X(z + 2\pi) = hX(z), \quad X(z + 2\pi\tau) = gX(z). \quad (6.324)$$

Clearly the condition $gh = hg$ is just the compatibility condition of these two equations. In a fancier language, the topological sectors of the M/G model quantized on the torus T^2 are classified by monodromy representations $\pi_1(T^2) \rightarrow G$, which correspond to conjugacy classes of commuting pairs of elements of G . Under $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$ the commuting pair (g, h) goes to $(g^a h^b, g^c h^d)$ hence the sum in the RHS of (6.323) is *naïvely* modular-invariant. As in Chap. 5 the *naïve* modular invariance may be spoiled by quantum *anomalous phases*. This cannot happen if the path integral in each (h, g) sector is left-right symmetric.⁴⁹ More generally, if the original model on M had a maximal local operator algebra $\mathfrak{A}_M$, after modding out G the orbifold operator algebra $\mathfrak{A}_X$ is also maximally local. The orbifold model is modular-invariant, hence consistent, iff $L_0 - \bar{L}_0 \in \mathbb{Z}$.

⁴⁹ For left-right *asymmetric* orbifolds, see [74].

Appendix 2: Classification of $c = 1$ CFTs

There is no rigorous proof, but the following is believed to be the complete list of CFTs with $c = 1$ [75]. Consider the compact scalar at the self-dual radius $R = \sqrt{2}$ (we set $\alpha' = 2$): this CFT is the $SU(2) \times SU(2)$ current algebra at level 1, i.e. the $SU(2)$ WZW model with $k = 1$ whose field takes value in the $SU(2)$ group manifold, cf. **Appendix 1** of Chap. 2. We can mod out (i.e. twist by) any finite subgroup $G \subset SU(2)$ acting by conjugacy $g \mapsto hgh^{-1}$, $g \in SU(2)$, $h \in G$. This action is not free but is symmetric between the left and the right $SU(2)$ currents, so such a twisting yields a symmetric orbifold. The finite subgroups of $SU(2)$ are classified by the McKay correspondence, cf. **Appendix 2** of Chap. 2. We denote a finite subgroup by the Cartan symbol of the corresponding finite-dimensional simply-laced Lie algebra. Each subgroup $G \subset SU(2)$ produces a $c = 1$ CFT, in fact a RCFT. Since -1 acts trivially, we consider only subgroups which contain -1 , and twist by the group $\hat{G} \equiv G/\{\pm 1\} \subset PSU(2)$ acting effectively.

Exercise 6.10 Show that $G = A_{2m-1} \simeq \mathbb{Z}_{2m}$ yields the compact scalar at radius $R = m\sqrt{2}$.

Exercise 6.11 Show that $G = D_m \simeq \mathbb{Z}_{2m} \rtimes \mathbb{Z}_2$ yields the $S^1/\mathbb{Z}_2$ model at radius $R = m\sqrt{2}$.

The above models have a marginal operator, corresponding to the changing the radius to an arbitrary real value R . In this way we get all S^1 and $S^1/\mathbb{Z}_2$ CFT at arbitrary *real* radius $R \geq \sqrt{2}$.

There remain 3 possibilities: the groups associated to E_6 , E_7 , E_8 , that is (respectively) the binary tetrahedral group, the binary octahedral group, and the binary icosahedral group. In this case there is a unique G -invariant primary of weights $(1, 1)$ i.e. $O(z, \bar{z}) \equiv \delta^{ij} j_i(z) \bar{j}_j(\bar{z})$. However this operator is not marginal since it does not satisfy condition (ii) at the end of Sect. 6.2.

Exercise 6.12 Show that $O(z, \bar{z}) O(w, \bar{w})$ contains $O(w, \bar{w})$ with a non-zero coefficient.

These 3 models are *rigid* (no conformal deformation) hence isolated in the space of $c = 1$ CFTs.

Exercise 6.13 Show that the partition functions of the non-Abelian orbifolds are:

$$Z(\tau)_{D_m} = \frac{1}{2} \left(Z(m\sqrt{2}, \tau)_{S^1} + 2Z(2\sqrt{2}, \tau)_{S^1} - Z(\sqrt{2}, \tau)_{S^1} \right) \quad (6.325)$$

$$Z(\tau)_{E_6} = \frac{1}{2} \left(2Z(3\sqrt{2}, \tau)_{S^1} + Z(2\sqrt{2}, \tau)_{S^1} - Z(\sqrt{2}, \tau)_{S^1} \right) \quad (6.326)$$

$$Z(\tau)_{E_7} = \frac{1}{2} \left(Z(4\sqrt{2}, \tau)_{S^1} + Z(3\sqrt{2}, \tau)_{S^1} + Z(2\sqrt{2}, \tau)_{S^1} - Z(\sqrt{2}, \tau)_{S^1} \right) \quad (6.327)$$

$$Z(\tau)_{E_8} = \frac{1}{2} \left(Z(5\sqrt{2}, \tau)_{S^1} + Z(3\sqrt{2}, \tau)_{S^1} + Z(2\sqrt{2}, \tau)_{S^1} - Z(\sqrt{2}, \tau)_{S^1} \right) \quad (6.328)$$

Thumb Rule: the Dynkin graph of the Lie algebra associated to a non-Abelian finite subgroup $G \subset SU(2)$ is a tree with 3 branches. Let ℓ_1, ℓ_2, ℓ_3 be the number of nodes in the branches (counting the central node). The partition function is

$$\frac{1}{2} \left(\sum_{i=1}^3 Z(\ell_i \sqrt{2}, \tau)_{S^1} - Z(\sqrt{2}, \tau)_{S^1} \right). \quad (6.329)$$

References

1. Kaluza, T.: Zum Unitätsproblem in der Physik, Sitzungsber. Preuss. Akad. Wiss. Berlin. (Math. Phys.) 966–972 (1921)
2. Klein, O.: Quantentheorie und fünfdimensionale Relativitätstheorie. Zeitschrift für Physik A **37**, 895–906 (1926)
3. Overduin, J.M., Wesson, P.S.: Kaluza-Klein gravity. Phys. Rep. **283**, 303–378 (1997). [arXiv:gr-qc/9805018](#)
4. Husemoller, D.: Fibre Bundles. In: Graduate Texts in Mathematics, vol. 20, 3rd edn. Springer, Berlin (1994)
5. Kobayashi, S., Nomizu, K.: Foundations of Differential Geometry, vols. 1 and 2. Wiley (1963)
6. Taubes, C.H.: Differential Geometry, Bundles, Connections, Metrics and Curvature. Oxford Graduate Texts in Mathematics, vol. 23. Oxford University Press (2011)
7. Milnor, J.W., Stasheff, J.D.: Characteristic Classes. Princeton University Press (1974)
8. Bott, R., Tu, L.W.: Differential Forms in Algebraic Topology. Graduate Texts in Mathematics, vol. 82. Springer, Berlin (1982)
9. Bröcker, T., Tom Dieck, T.: Representations of Compact Lie Groups. Springer, Berlin (1985)
10. Iwaniec, H., Kowalski, E.: Analytic Number Theory. AMS Colloquium Publications, vol. 53. AMS (2004)
11. Carlson, J., Müller-Stach, S., Peters, C.: Period Mappings and Period Domains. Cambridge studies in advanced mathematics, vol. 168, 2nd edn. Cambridge University Press (2017)
12. Deligne, P., de Griffiths, T.: Séminaire Bourbaki Exp. Springer Lecture Notes in Mathematics, vols. 376, 180. Springer, Berlin (1970)
13. Griffiths, P.: Topics in Transcendental Algebraic Geometry. Princeton University Press (1984)
14. Helgason, S.: Differential Geometry, Lie Groups and Symmetric Spaces. Academic Press (1978)
15. Hatcher, A.: Algebraic Topology. Cambridge University Press (2002)
16. Huybrechts, D.: Complex Geometry: An Introduction. Universitext. Springer, Berlin (2005)
17. Dijkgraaf, R., Verlinde, E., Verlinde, H.: $c = 1$ conformal field theories on Riemann surfaces. Comm. Math. Phys. **115**, 649–690 (1988)
18. Zamolodchikov, A.B.: Irreversibility of the flux of the renormalization group in 2D field theory. JETP Lett. **43**, 730–732 (1986)
19. Kadanoff, L.P.: Multicritical behaviour at the Kosterlitz-Thouless critical point. Ann. Phys. **120**, 39 (1979)
20. Mumford, D.: Tata Lectures on Theta. Modern Birkhäuser Classics, vol. I, II, III. Springer, Berlin (2007)
21. Giverson, A., Porrati, M., Rabinovici, E.: Target space duality in string theory. Phys. Rept. **244**, 77–202 (1994). [arXiv:hep-th/9401139](#)
22. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](#)
23. Griffiths, P., Harris, J.: Principles of Algebraic Geometry. Wiley (1978)
24. Komargodski, Z., Schwimmer, A.: On renormalization group flows in four dimensions. JHEP **12**, 099 (2011). [arXiv:1107.3987](#)
25. Ooguri, H., Vafa, C.: On the geometry of the string landscape and the swampland. Nucl. Phys. B **766**, 21–33 (2007). [arXiv:hep-th/0605264](#)
26. Banks, T., Seiberg, N.: Symmetries and strings in field theory and gravity. Phys. Rev. D **83**, 084019 (2011). [arXiv:1011.5120](#)
27. Cecotti, S.: Supersymmetric Field Theories. Geometric Structures and Dualities. Cambridge University Press (2015)
28. Buscher, T.H.: A symmetry of the string background field equations. Phys. Lett. B **194**, 59–62 (1987)
29. Buscher, T.H.: Path integral derivation of quantum duality in nonlinear sigma models. Phys. Lett. B **201**, 466–472 (1988)

30. Serre, J.-P.: *A Course in Arithmetics*. Springer, Berlin (1973)
31. Conway, J.H., Sloane, N.J.A.: *Sphere Packings, Lattices and Groups*, 3rd edn. Springer, Berlin (1999)
32. Knapp, A.W.: *Lie Groups Beyond an Introduction*. Birkhäuser (1996)
33. Bump, D.: *Lie Groups*. Graduate Text in Mathematics, vol. 225. Springer, Berlin (2004)
34. Soulé, C.: An introduction to arithmetic groups. In: *Lectures at the Conference. Frontiers in Number Theory, Physics and Geometry*, Les Houches, 9–12 Mar 2003. [arXiv:math/0403390](https://arxiv.org/abs/math/0403390)
35. Postnikov, M.M.: *Geometry VI. Riemannian Geometry*. Encyclopaedia Mathematica Sciences, vol. 91. Springer, Berlin (2001)
36. Besse, A.: *Einstein Manifolds*. Springer, Berlin (1987)
37. Borel, A.: *Introduction Aux Groupes Arithmétiques*. Hermann (1969)
38. Morris, D.W.: *Introduction to Arithmetic Groups*. [arXiv:math/0106063](https://arxiv.org/abs/math/0106063)
39. Serre, J.-P.: Bounds for the orders of the finite subgroups of $G(k)$. In: Geck, M., Testerman, D., Thévenaz, J. (eds.) *Group Representation Theory*. EPFL Press (2006)
40. Ji, L.: *Arithmetic Groups and Their Generalizations. What, Why, and How*. AMS/IP Studies in Advanced Mathematics, vol. 43. AMS (2008)
41. Borel, A.: Density and maximality of arithmetic subgroups. *J. Reine Angew. Math.* **244**, 78–89 (1966)
42. Langlands, R.P.: Volume of the fundamental domain for some arithmetical subgroup of Chevalley groups. In: *Algebraic Groups and Discontinuous Subgroups, Proceedings of Symposia in Pure Mathematics*, vol. IX, pp. 235–257. AMS (1966)
43. Bourbaki, N.: *Éléments de mathématique. Groupes et algèbres de Lie. Chapitre 4 à 6*. Springer, Berlin (2007)
44. Macdonald, I.G.: The volume of a compact Lie group. *Inven. math.* **56**, 93–96 (1980)
45. Dixon, L.J., Harvey, J.A., Vafa, C., Witten, E.: Strings on orbifolds. *Nucl. Phys. B* **261**, 678–686 (1985)
46. Dixon, L.J., Harvey, J.A., Vafa, C., Witten, E.: Strings on orbifolds (II). *Nucl. Phys. B* **274**, 285–314 (1986)
47. Lando, S.K., Zvonki, A.K.: *Graphs on surfaces and their applications*. Encyclopaedia of Mathematical Sciences (EMS), vol. 141. Springer, Berlin (2004)
48. SGA1. *Revêtements étales et groupe fondamental*. Documents Mathématiques (Paris) 3. Société Mathématique de France, Paris (2003); Séminaire de géométrie algébrique du Bois Marie 1960–1961, directed by A. Grothendieck. With two papers by M. Raynaud, Updated and annotated reprint of the 1971 original. *Lecture Notes in Mathematics*, vol. 224. Springer, Berlin (2003)
49. Farkas, H.M., Kra, I.: *Riemann Surfaces*. Graduate Texts in Mathematics, vol. 71. Springer, Berlin (1992)
50. Al, B.: Zamolodchikov, conformal scalar field on the hyperelliptic curve and critical Ashkin-Teller multipoint correlation functions. *Nucl. Phys. B* **285**(FS19), 481–503 (1987)
51. Elitzur, S., Gross, E., Rabinovici, E., Seiberg, N.: Aspects of bosonization in string theory. *Nucl. Phys. B* **283**, 413–432 (1987)
52. Yang, S.K.: Modular invariant partition function of the Ashkin-Teller model on the critical line and $N = 2$ superconformal invariance. *Nucl. Phys. B* **285**(FS19), 183–203 (1987)
53. Saleur, H.: Partition function of the two-dimensional Ashkin-Teller model on the critical line. *J. Phys. A* **20**, L1127–L1133 (1987)
54. Vafa, C.: Modular invariance and discrete torsion on orbifolds. *Nucl. Phys. B* **273**, 592 (1986)
55. Vafa, C., Witten, E.: On orbifolds with discrete torsion. *J. Geom. Phys.* **15**, 189–214 (1995). [arXiv:hep-th/9409188](https://arxiv.org/abs/hep-th/9409188)
56. Brown, K.S.: *Cohomology of Groups*. Graduate Text in Mathematics, vol. 87. Springer, Berlin (1982)
57. Polchinski, J.: TASI lectures on D-branes. In: Efthimiou, C., Greene, B. (eds.) *Fields, Strings, and Duality*, TASI 1996, pp. 293–356. World-Scientific (1997) [arXiv:hep-th/9611050](https://arxiv.org/abs/hep-th/9611050)
58. Johnson, C.V.: *D-Branes*. Cambridge University Press (2003)

- 59. Sen, A.: Tachyon condensation on the brane antibrane system. JHEP **8**, 012 (1998). [arXiv:hep-th/9805170](#)
- 60. Seiberg, N., Witten, E.: String theory and noncommutative geometry. JHEP **09**, 032 (1999). [arXiv:hep-th/9908142](#)
- 61. Born, M., Infeld, L.: Foundations of the new field theory. Proc. Royal Soc. A **144**, 425–451 (1938)
- 62. Tseytlin, A.A.: Born-Infeld action, supersymmetry and string theory. In: The many faces of the superworld, pp. 417–452 (2000). [arXiv:hep-th/9908105](#)
- 63. Bialynicki-Birula, I.: In: Jancewicz, B., Lukierski, J. (eds.) Festschrift of J. Lopuszanski, Quantum Theory of Particles and Fields, pp. 31–42. World Scientific, Singapore (1983)
- 64. Polchinski, J.: Dirichlet branes and Ramond-Ramond charges. Phys. Rev. Lett. **75**, 4724–4727 (1995). [arXiv:hep-th/9510017](#)
- 65. Polchinski, J.: String Theory, vols. 1, 2. Cambridge University Press, Cambridge (2003)
- 66. Dai, J., Leigh, R.G., Polchinski, J.: New connections between string theories. Mod. Phys. Lett. A **4**, 2073–2083 (1989)
- 67. Sagnotti, A.: Non-perturbative Quantum Field Theory. In: Mack, G., et al. (eds.), p. 521. Pergamon Press (1988)
- 68. Govaerts, J.: Quantum consistency of open string theories. Phys. Lett. B **220**, 77–86 (1989)
- 69. Ishibashi, N., Onogi, T.: Open string model building. Nucl. Phys. B **318**, 239–280 (1989)
- 70. Pradisi, G., Sagnotti, A.: Open string orbifolds. Phys. Lett. B **216**, 59 (1989)
- 71. Horava, P.: Strings on world sheet orbifolds. Nucl. Phys. B **327**, 461–48 (1989)
- 72. Sagnotti, A.: Closed strings and their open string descendants. Phys. Rept. **184**, 167–175 (1989)
- 73. Dabholkar, A.: Lectures on Orientifolds and Duality. [arXiv:hep-th/9804208](#)
- 74. Narain, K.S., Sarmadi, M.H., Vafa, C.: Asymmetric orbifolds. Nucl. Phys. B **288**, 551–577 (1987)
- 75. Ginsparg, P.: Curiosities at $c = 1$. Nucl. Phys. B **295**(FS21), 153–170 (1988)

Chapter 7

The Heterotic String



Abstract In this chapter we construct and study a new class of string theories: the *heterotic* ones which on the left side are the bosonic string and on the right one the superstring. Two of these theories are supersymmetric (16 supercharges) with spacetime gauge group $Spin(32)$ and $E_8 \times E_8$, respectively. We construct these theories in two different (but equivalent) ways: using either chiral fermions or chiral bosons to represent their ‘internal’ left-moving 2d d.o.f. We study the toroidal compactifications of these theories, their dualities, and equivalences. We also discuss the non-SUSY heterotic strings, and in particular the tachyon-free one with gauge group $SO(16) \times SO(16)$. Finally we describe the BPS states of the SUSY heterotic strings. In the **Appendix** the $\mathcal{N} = 2$ superstring is briefly reviewed.

7.1 Constructing String Models

In Chap. 5 we constructed three interesting supersymmetric string models, i.e. Type IIA, Type IIB, and Type I with $G = SO(32)$ as a straightforward supersymmetric version of the bosonic string. These models do not exhaust the list of consistent (and interesting) string theories. In this chapter we attempt to be more systematic in our quest for consistent string theories.

Both the bosonic string and the superstring were constructed from the point of view of the world-sheet theories in terms of an algebra $\mathcal{A}_{\text{con}}$ of gauge constraints, implemented *à la* OCQ or (more modernly) *à la* BRST. The algebra of constraints was the algebra of modes of a set of (anti-)holomorphic currents of various spins: the energy-momentum tensor $T(z)$ in the bosonic case, and the energy-momentum together with its fermionic superpartner $T_F(z)$ (the $\mathcal{N} = 1$ supercurrent) in the superstring. Looking for a more systematic generalization of our previous constructions, the first question we have to address is the classification of the possible chiral symmetry algebras $\mathcal{A}_{\text{con}}$ which may be used as an algebra of gauge constraints to define the world-sheet theory of a meaningful string model. This analysis was already performed in Sect. 2.10, and its implications for the construction of “new” string theories were discussed at the end of that section. We recall the conclusion: under the natural

assumptions in Sect. 2.10, we have only 3 candidate constraint chiral algebras for string theories:

- the Virasoro algebra (as in the bosonic string)
- the $\mathcal{N} = 1$ superconformal algebra (as in the superstring)
- the $\mathcal{N} = 2$ superconformal algebra.

The third possibility is analyzed in the **Appendix**: we get a consistent quantum system which has great mathematical interest, but is not viable as a physical theory for two reasons: (i) the target space-time *cannot have* Lorentzian signature (1 time and $d - 1$ spatial directions), and (ii) the spectrum consists only of massless scalars. The $\mathcal{N} = 2$ superstring has no propagating graviton, so it is *not* a Quantum Gravity.¹

There is, however, yet another possibility [1, 2]: $\mathcal{A}_{\text{con}}$ is the chiral algebra of left-moving constraints; of course, there is also a right-moving such algebra $\tilde{\mathcal{A}}_{\text{con}}$. The two chiral algebras *need not* be isomorphic, as they were in the bosonic string and the superstring. We may take as algebra of constraints a pair of superconformal algebras $(\mathcal{A}_{\text{con}}, \tilde{\mathcal{A}}_{\text{con}})$ with $\mathcal{N} \neq \tilde{\mathcal{N}}$, say $(\mathcal{N}, \tilde{\mathcal{N}}) = (0, 1)$ or equivalently $(1, 0)$. This asymmetric possibility leads to the *heterotic string* which is the topic of this chapter.

One basic aspect should be always kept in mind: heterotic strings, having $\mathcal{N} \neq \tilde{\mathcal{N}}$ local SUSY on the world-sheet, *are necessarily closed and oriented*, since boundary conditions and Ω -projections identify the left- and right-moving SCFTs and this makes sense only if they are isomorphic.

These conclusions hold *only* under the assumptions in Sect. 2.10. There are *several* interesting constructions, which do not obey those conditions, to which our conclusions do not apply. We already mentioned one example, the *topological string theories* [3, 4]. Another important example is the *Green-Schwarz formulation of the superstring* [5] which leads to a physically equivalent theory but starting from a different world-sheet formulation in which target space supersymmetry is manifest from the start without the “ad hoc” GSO projection. Unfortunately this version of the theory is difficult to quantize, and less friendly as a computational tool, since in a covariant gauge the world-sheet QFT is *not free* even for superstrings in flat space.

7.2 The $SO(32)$ and $E_8 \times E_8$ Heterotic Strings in 10d

The $(0, 1)$ heterotic string [1] combines the constraints and ghosts of the bosonic string in the left-moving side with the constraints and ghosts of the Type II superstring in the right-moving side. The geometric interpretation of the target space M as physical spacetime requires the embedding coordinates X^μ of the world-sheet Σ into M to be ordinary (i.e. non-chiral) scalar fields. Hence the maximal number of space-time dimensions d in which the heterotic string may move is 10 from the critical dimension of its SUSY side. For now we focus on $d = 10$; the case $d < 10$

¹ In the *strict* sense that it does not contain Einstein gravity. The $\mathcal{N} = 2$ string *is* a Quantum Gravity in a broader sense.

will be discussed later in the chapter. The embedding fields with their right-moving $(0, 1)$ superpartners

$$X^\mu(z, \bar{z}), \quad \tilde{\psi}^\mu(\bar{z}), \quad \mu = 0, 1, \dots, 9, \quad (7.1)$$

have total left/right Virasoro central charges $(c, \tilde{c}) = (10, 15)$. We know from table (2.541) that the central charges of the ghosts are

$$(c_{\text{ghost}}, \tilde{c}_{\text{ghost}}) = (-26, -15), \quad (7.2)$$

so, to guarantee the nilpotency of the BRST charge, we have to add conformal matter with central charges $(c, \tilde{c}) = (16, 0)$. The simplest possibility is to take 32 real free left-moving spin- $\frac{1}{2}$ fields

$$\lambda^A(z), \quad A = 1, 2, \dots, 32. \quad (7.3)$$

In the superconformal gauge the total matter action reads²

$$S_{\text{matt}} = \frac{1}{4\pi} \int d^2z \left(\partial X^\mu \bar{\partial} X_\mu + \lambda^A \bar{\partial} \lambda^A + \tilde{\psi}^\mu \partial \tilde{\psi}_\mu \right), \quad (7.4)$$

which leads to the free OPEs

$$X^\mu(z, \bar{z}) X^\nu(0, 0) \sim -\eta^{\mu\nu} \log |z|^2, \quad (7.5)$$

$$\lambda^A(z) \lambda^B(0) \sim \frac{\delta^{AB}}{z}, \quad (7.6)$$

$$\tilde{\psi}^\mu(\bar{z}) \tilde{\psi}^\nu(0) \sim \frac{\eta^{\mu\nu}}{\bar{z}}. \quad (7.7)$$

The matter energy-momentum tensors and supercurrent are

$$T_B = -\frac{1}{2} \partial X^\mu \partial X_\mu - \frac{1}{2} \lambda^A \partial \lambda^A, \quad (7.8)$$

$$\tilde{T}_B = -\frac{1}{2} \bar{\partial} X^\mu \bar{\partial} X_\mu - \frac{1}{2} \tilde{\psi}^\mu \bar{\partial} \tilde{\psi}_\mu, \quad (7.9)$$

$$\tilde{T}_F = i \tilde{\psi}^\mu \bar{\partial} X_\mu. \quad (7.10)$$

The world-sheet matter theory has a $SO(9, 1) \times O(32)$ symmetry. $SO(9, 1)$ is the target space Lorentz group. The $O(32)$ acting on the λ^A is an *internal* symmetry. No λ^A can have a time-like signature in a unitary theory because on the left side there is no fermionic constraint to get rid of fermionic negative-norm states. So, while the world-sheet action of the λ^A looks similar to action for ψ^μ 's in the superstring, their physical implications are quite different because of the different gauge constraints.

² We choose our units so that $\alpha' = 2$.

The right-moving ghosts and constraints are as in the superstring, and the left-moving ones are as in the bosonic string. Therefore the left- and right-moving BRST currents are the same ones as in the bosonic and superstring, respectively, and they are well-defined as long as the fermionic periodic b.c. are consistent with BRST invariance, i.e. the same ones for all $\tilde{\psi}^\mu$ and the $\tilde{\beta}$, $\tilde{\gamma}$ ghosts. The BRST charges are nilpotent since the total central charges

$$c_{\text{tot}} = c_{\text{matter}} + c_{\text{ghost}} \quad \tilde{c}_{\text{tot}} = \tilde{c}_{\text{matter}} + \tilde{c}_{\text{ghost}} \quad (7.11)$$

vanish. The proof of the no-ghost theorem for the bosonic string and the superstring worked independently for left- and right-movers, so they apply without modification in the heterotic set-up, which is then free of space-time ghosts, that is, unitary. All these results continue to hold if we replace the spatial part of the free matter theory (7.4) by any *unitary* SCFT (0,1) with equivalent Virasoro central charges.

To complete the definition of the theory, we need to specify which sectors we keep in the spectrum. The general considerations in Sect. 5.1, 5.2 still apply: the world-sheet operator algebra $\mathfrak{A}$ should be maximal local with integral spins. The analysis of all $\mathfrak{A}$'s satisfying these conditions is more complicated than in Type II, because no principle requires all λ^A 's to have the same b.c. Periodicity of $T_B(w)$ only requires that the λ^A are periodic up to an arbitrary $O(32)$ rotation

$$\lambda^A(w + 2\pi) = O^{AB} \lambda^B(w). \quad (7.12)$$

Nine 10-dimensional theories based on the world-sheet action (7.4) exist:

- *six* of them have tachyons, and so are consistent only in the sense the bosonic string is consistent;
- *two* have space-time supersymmetry (hence tadpole- and tachyon-free);
- *one* is tachyon-free but has no space-time SUSY.

In this section we focus on the two supersymmetric theories. The other 10d heterotic strings will be discussed in Sect. 7.3.

Space-Time SUSY Versus Right GSO Projection

In the Type IIA and Type IIB superstring the GSO projection acted separately on the left- and right-movers. This remains true in all *supersymmetric* heterotic theory. We revisit the argument (already presented in Sect. 3.8). The world-sheet current associated to space-time supersymmetry (in, say, the $-\frac{1}{2}$ picture) is

$$\tilde{Q}_\alpha(\bar{z}) = \tilde{S}_\alpha(\bar{z}) e^{-\tilde{\phi}(\bar{z})/2}, \quad (7.13)$$

cf. Sect. 3.8 in particular Eq. (3.284). $\tilde{Q}_\alpha$ with (say) chirality +1 transforms in the $\mathbf{16}_s$ of $Spin(9, 1)$ which we Wick rotate to $Spin(10)$. In the sense of Sect. 3.1, the current $\tilde{Q}_\alpha(z)$ has a $Spin(10, 2)$ weight of the form

$$\left(\tilde{w}, -\frac{1}{2}\right) \quad \text{with } \tilde{w} \text{ a weight of the } \mathbf{16}_s \text{ of } Spin(10). \quad (7.14)$$

The heterotic string is supersymmetric in space-time iff the action of the *supercharge*

$$Q_\alpha \equiv \oint \frac{d\bar{z}}{2\pi i} \tilde{Q}_\alpha(\bar{z}), \quad (7.15)$$

is well-defined on the physical vertices, so that the SUSY Ward identities hold for all amplitudes, as demonstrated in Sect. 3.8.1. The action of Q_α is well-defined iff the OPE of the current $\tilde{Q}_\alpha(\bar{z})$ with all vertex operators $\tilde{O}(\bar{z})$ is *single-valued*. Let $\tilde{\lambda} \equiv (\tilde{\pi}, \tilde{q})$ be the $Spin(10, 2)$ weight of a right-moving operator $\tilde{O}(\bar{z})$, $\tilde{\pi}$ being its $Spin(10)$ weight and $\tilde{q}$ its (right-moving) picture charge. $\tilde{O}$ is local with respect to $\tilde{Q}_\alpha(\bar{z})$ iff (cf. Sect. 3.1)

$$\tilde{\pi} \cdot \tilde{w} + \frac{\tilde{q}}{2} \in \mathbb{Z} \quad \text{for all weights } \tilde{w} \text{ of the } \mathbf{16}_s. \quad (7.16)$$

We already know that this condition is equivalent to

$$\sum_{i=1}^6 \tilde{\lambda}_i \equiv \sum_{i=1}^5 \tilde{\pi}_i + \tilde{q} = 0 \pmod{2}, \quad (7.17)$$

which in turn is equivalent to $[(-1)^{\tilde{F}}, \tilde{O}] = 0$ for all vertex operators $\tilde{O}$ of physical states i.e. to the GSO projection. In other words:

in all maximally local operator algebra $\mathfrak{A}$ containing (7.13) the right-moving Fermi d.o.f. are GSO projected

We write $\mathfrak{A}_R$ for the GSO-projected algebra of the right-moving spinor d.o.f. (cf. Sect. 3.1). The partition function of the right-moving spinors $\tilde{\psi}^\mu, \tilde{\beta}, \tilde{\gamma}$ is as in Type II superstrings, that is, $Z_\psi^+(\tau)^*$, cf. Eqs. (5.102), (5.105). From Eq. (5.106) we have

$$Z_\psi^+(-1/\tau)^* = Z_\psi^+(\tau)^*, \quad Z_\psi^+(\tau + 1)^* = e^{-2\pi i/3} Z_\psi^+(\tau)^*. \quad (7.18)$$

$SO(32)$ Heterotic String

To get a consistent string theory we should impose the consistency requirements in Sect. 5.2, that is, the world-sheet physical operator algebra $\mathfrak{A}$ should be maximally local and spin-integral (i.e. $h - \bar{h} \in \mathbb{Z}$). Spacetime supersymmetry fixes the right-moving spinor part of $\mathfrak{A}$ to be $\mathfrak{A}_R$, so that

$$\mathfrak{A} = \mathfrak{A}_X \otimes \mathfrak{A}_{\text{gh}} \otimes \mathfrak{A}_L \otimes \mathfrak{A}_R, \quad (7.19)$$

where $\mathfrak{A}_X$, $\mathfrak{A}_{\text{gh}}$, and $\mathfrak{A}_L$ are, respectively, the OPE algebras of the X^μ , the $\mathbf{Diff}^+$ ghosts, and the left-moving Fermi d.o.f. $\mathfrak{A}$ is maximally local spin-integral iff $\mathfrak{A}_L$ is.

The $SO(8n)$ current algebra of a set λ^A , $A = 1, \dots, 8n$ of free fermions was studied in Sect. 2.9.2: that analysis applies to the left-movers λ^A of the heterotic string.

GSO Projections on the Left-Movers A procedure which is guaranteed to produce a maximally local spin-integral OPE algebra $\mathfrak{A}_L$ is to perform the GSO projection as in Sect. 2.9.2. One takes the *same* sign periodicity for all 32 fermions

$$\lambda^A(w + 2\pi) = -(-1)^\alpha \lambda^A(w) \quad \text{for all } A, \quad \alpha = 0, 1, \quad (7.20)$$

and impose the chirality projection

$$(-1)^F = +1 \quad (7.21)$$

where F is the left-moving fermion number mod 2. Bosonization, vertices, *etc.* work exactly as in Sect. 2.9.2, and shall not be repeated here.

Modular Invariance The GSO-projected algebra of the left-moving fermions, $\mathfrak{A}_L$, is maximally local and spin-integral with central charges $(c, \tilde{c}) = (16, 0)$: its partition function $Z_{16}(\tau)$ transforms under $SL(2, \mathbb{Z})$ as predicted by Eq. (5.36)

$$Z_{16}(-1/\tau) = Z_{16}(\tau), \quad Z_{16}(\tau + 1) = e^{2\pi i/3} Z_{16}(\tau). \quad (7.22)$$

We check this prediction by computing $Z_{16}(\tau)$ explicitly. In terms of the functions $Z_\beta^\alpha(\tau)$ defined and studied in Sect. 5.4, the partition function of the λ^A 's is

$$\begin{aligned} Z_{16}(\tau) &= \frac{1}{2} \sum_{\alpha, \beta=0}^1 \text{Tr}_\alpha [(-1)^{\beta F} q^{L_0 - c/24}] = \\ &= \frac{1}{2} [Z_0^0(\tau)^{16} + Z_1^0(\tau)^{16} + Z_0^1(\tau)^{16} + Z_1^1(\tau)^{16}]. \end{aligned} \quad (7.23)$$

Using formulae in BOX 5.6, we see that the modular transformation $S: \tau \mapsto -1/\tau$ permutes the four functions $Z_\beta^\alpha(\tau)^{16}$ in the large bracket, while $T: \tau \mapsto \tau + 1$ acts on them as a permutation together with multiplication by $(e^{2\pi i/3})^4 \equiv e^{2\pi i/3}$ in full agreement with (7.22). Comparing with (7.18), we see that the partition function of *all* 2d spinor fields, left- and right-moving,

$$Z_{16}(\tau) Z_\psi^+(\tau)^* \quad (7.24)$$

is *fully modular invariant*. The partition function $Z_{16}(\tau)$ in (7.23) has the same modular properties as the function $Z_\psi^+(\tau)$ in the Type II string, but now all terms in the sum have sign $+$. If we interpret $Z_\psi^+(\tau)^*$ as a “twisted” partition function for the CFT of the 8 transverse fermions $\tilde{\psi}^i$ (as we did in Chap. 5), the varying signs reflect the difference in modular properties between a maximal local algebra with $h - \tilde{h} \in \mathbb{Z}$ and one with $h - \tilde{h} \in \frac{1}{2}\mathbb{Z}$ as discussed at the end of Sect. 5.2. In the more intrinsic interpretation of $Z_\psi^+(\tau)^*$ as the partition function of the full

GSO-projected $\tilde{\psi}^\mu$, $\tilde{\beta}$, $\tilde{\gamma}$ system, which is maximal local and spin integral (cf. Sect. 3.1) the minus signs arise from the ghosts' zero-modes. The different signs can be understood physically in several ways. First of all, passing from 8 to 32 fermions, all signs in the modular transformations get raised to the fourth power, and minuses become pluses: the first three terms in the RHS of (7.23) now have the same sign. Let us recall the physical origin of the various minus signs³ in the expression for $Z_\psi^+(\tau)$ and compare with the different situation for the left-moving λ^A 's in the heterotic string. (i) The relative minus sign between the first and second terms of $Z_\psi^+(\tau)$ arises from $(-1)^{F_{\text{GSO}}} = (-1)^{F_\psi} (-1)^{F_{\beta,\gamma}}$ and the fact that in the standard picture⁴ the ghosts' Fermi parity in the NS sector is $(-1)^{F_{\beta,\gamma}} = -1$. In the left-moving sector of the heterotic string we have no β, γ ghosts, hence no extra sign. (ii) the relative sign between the first and third terms in $Z_\psi^+(\tau)$ comes from the spacetime Fermi statistics. The λ^A are space-time scalars, and so are their R sector states, which then should be bosonic.⁵

Thus modular invariance, conservation of $(-1)^F$ in OPE, and space-time Spin & Statistics are all consistent with the partition function (7.23).

Massless States The right sides of Type II and heterotic strings are identical. There is no tachyon, and the right-moving massless level contains a vector in the $\mathbf{8}_v$ of the transverse $SO(8)$, and a (on-shell) Majorana-Weyl spinor in the $\mathbf{8}_s$ of $SO(8)$.

In the left-moving side, the $(\text{mass})^2/2$ is given by $N + a$, where N is the transverse oscillator level (whose spectrum is $\frac{1}{2}\mathbb{N}$) while the “normal-order” constant a is

$$\text{NS:} \quad -\frac{1}{24} \overbrace{\left(8 + \frac{32}{2}\right)}^{c_{\chi^i} + c_\lambda} \equiv \overbrace{-1}^{h_c}, \quad (7.25)$$

$$\text{R:} \quad \overbrace{\frac{16}{8}}^{h_s} - \frac{1}{24} \overbrace{\left(8 + \frac{32}{2}\right)}^{c_{\chi^i} + c_\lambda} \equiv \overbrace{+1}^{h_s + h_c} \quad (7.26)$$

where c_{χ^i} (resp. c_λ) is the central charge of the transverse scalars (resp. of λ^A), h_c is the weight of the reparametrization ghost c , and h_s the weight of the $SO(32)$ spin field. In the above equations the expression before $\equiv$ stands for the ‘old’ evaluation of a and the one after $\equiv$ for the ‘modern’ BRST one (they agree of course).

³ The power of Z_1^1 transforms into itself, and its sign depends on the chirality in the R sector. The theories obtained by flipping the sign of the chirality in the R sector are physically equivalent.

⁴ This is the picture most directly related to the ‘transverse’ picture of the physical states that we used in Chap. 5 to write the superstring partition function in terms of the fermionic ones.

⁵ The proof of the *10d Spin & Statistics Theorem* in BOX 5.3 applies, with no modification, to the heterotic case too. As demonstrated there, the Grassmann $\mathbb{Z}_2$ degree of a physical state is $2(q - \bar{q}) \bmod 2$. In the heterotic string $q \equiv 0$ and hence the spacetime statistics of the left-moving NS and R states are both even. The correct statistics (and hence correct signs in one-loop amplitudes) is enforced by the ghosts' zero-modes, as evident from the full GSO-projected $\tilde{\psi}^\mu$, $\tilde{\beta}$, $\tilde{\gamma}$ perspective.

Table 7.1 Low-lying states in the heterotic string

m^2	NS	R	$\widetilde{\text{NS}}$	$\widetilde{\text{R}}$
-2	(1, 1)	–	–	–
0	(8_v , 1) $\oplus$ (1 , 496)	–	8_v	8_s

The left-moving ground state is therefore a tachyon. The first excited states

$$\lambda_{-1/2}^A |0; k\rangle_{\text{NS}}, \quad (7.27)$$

have $N + a = -\frac{1}{2}$, but this would-be tachyon is eliminated by the chirality projection⁶ (7.21). Left-moving massless states can be obtained in two ways:

$$\alpha_{-1}^i |0; k\rangle_{\text{NS}}, \quad \lambda_{-1/2}^A \lambda_{-1/2}^B |0; k\rangle_{\text{NS}}. \quad (7.28)$$

The λ^A transform in the **32** of the $SO(32)$ internal symmetry. Under the full symmetry $SO(8) \times SO(32)$, the left-moving NS ground state is invariant, (**1**, **1**). The second state in (7.28) is antisymmetric in $A \leftrightarrow B$, so the states (7.28) transform as

$$(\mathbf{8}_v, \mathbf{1}) \oplus (\mathbf{1}, \Lambda_2). \quad (7.29)$$

The 2-index antisymmetric representation of $SO(32)$, Λ_2 (the adjoint), has dimension

$$\frac{1}{2}(32 \times 31) = \mathbf{496} = 2 \times \mathbf{248}. \quad (7.30)$$

Table 7.1 shows the low-lying states in both sides of the heterotic string. Left-movers are listed with their $SO(8) \times SO(32)$ quantum numbers, and right-movers with their $SO(8)$ quantum numbers. Closed string states combine left- and right-moving states *with the same mass*. The left sector, like the bosonic string, has a would-be tachyon, but there is no matching right-moving tachyon, so the theory is tachyon-free.⁷ At the massless level the left-right tensor product

$$(\mathbf{8}_v, \mathbf{1}) \otimes (\mathbf{8}_v \oplus \mathbf{8}_s) = \overbrace{(\mathbf{1}, \mathbf{1}) \oplus (\mathbf{28}, \mathbf{1}) \oplus (\mathbf{35}, \mathbf{1})}^{\text{bosons}} \oplus \overbrace{(\mathbf{56}_s, \mathbf{1}) \oplus (\mathbf{8}_c, \mathbf{1})}^{\text{fermions}}, \quad (7.31)$$

produces the Type I supergravity multiplet⁸ consisting of the spacetime metric $G_{\mu\nu}$, a 2-form gauge field $B_{\mu\nu}$, a dilaton Φ , as well as a MW gravitino ψ_μ and a MW dilatino χ of opposite chirality. The other left-right tensor product

⁶ The ground state has $(-1)^F = +1$, since it is the fermion parity of the actual SL_2 -invariant vacuum not—as it was in Type II—of the $|q = -1\rangle$ Bose sea.

⁷ This was expected since, by its very construction, the theory is supersymmetric in spacetime.

⁸ Compare with the corresponding sector in the Type I superstring of Chap. 5.

$$(\mathbf{1}, \mathbf{496}) \otimes (\mathbf{8}_v \oplus \mathbf{8}_s) = \overbrace{(\mathbf{8}_v, \mathbf{496})}^{\text{gauge vectors}} \oplus \overbrace{(\mathbf{8}_s, \mathbf{496})}^{\text{gauginos}}, \quad (7.32)$$

is the 10d $\mathcal{N} = 1$ gauge supermultiplet in the adjoint of $SO(32)$. The world-sheet global symmetry $SO(32)$ is promoted to a spacetime *gauge* symmetry, as expected.

This is the same massless content we found in the Type I $SO(32)$ string. Do we have constructed a *new* theory, or just found a different construction of the same physical system? The two theories have the same massless sector (and hence the same low-energy physics) *but* different massive spectra. In open string theory the $SO(32)$ gauge quantum numbers are carried by a vector $SO(32)$ index at each endpoint, so even at the massive level the $SO(32)$ representations which appear are all subrepresentations of the 2-index tensor. On the contrary, in the heterotic string the gauge quantum numbers are carried by the field λ^A which propagate in the bulk of the world-sheet. At the massive level *any* number of fermions can be excited, producing arbitrarily large representations of the spacetime gauge group $Spin(32)$; in particular, states in the left-moving R-sector, are $Spin(32)$ spinors. Thus *perturbatively* these two $SO(32)$ strings look quite different. We shall see in Chap. 13 that *non-perturbatively* they are ‘the same’ theory.

The $E_8 \times E_8$ Heterotic String

The second supersymmetric 10d heterotic string is obtained by dividing the 32 λ^A into two sets of 16 with independent boundary conditions

$$\lambda^A(w + 2\pi) = \begin{cases} \eta \lambda^A(w) & A = 1, 2, \dots, 16 \\ \eta' \lambda^A(w) & A = 17, 18, \dots, 32, \end{cases} \quad (7.33)$$

where η and η' take values ± 1 . Dually, there are two chirality operators

$$(-1)^F, \quad (-1)^{F'}, \quad (7.34)$$

which anticommute with for $A = 1, \dots, 16$, and respectively $A = 17, \dots, 32$, and commute with the other λ^A . The two sets of λ^A generate a $SO(16) \times SO(16)$ current algebra to which we may apply the results of Sect. 2.9.2. We take the separated GSO projections on each left-moving $SO(16)$ current algebra as in Sect. 2.9.2. In addition, to get a SUSY theory, we perform the standard GSO projection in the right-sector as before. This means that in the partition function we have to sum over the $2^3 = 8$ distinct periodicity sectors while inserting the three projections

$$(-1)^F = (-1)^{F'} = (-1)^{\tilde{F}} = +1. \quad (7.35)$$

Modular Invariance The left-moving algebra $\mathfrak{A}_L$ is the tensor product of two copies of the GSO-projected algebra $\mathfrak{A}_{16}$ of 16 left-moving fermions. $\mathfrak{A}_{16}$ is maximally local and spin-integral with $(c, \tilde{c}) = (8, 0)$, so its partition function $Z_8(\tau)$ must satisfy

$$Z_8(-1/\tau) = Z_8(\tau), \quad Z_8(\tau + 1) = e^{4\pi i/3} Z_8(\tau), \quad (7.36)$$

by Eq. (5.36). This prediction is confirmed by its explicit expression

$$Z_8(\tau) = \frac{1}{2} \left(Z_0^0(\tau)^8 + Z_1^0(\tau)^8 + Z_0^1(\tau)^8 + Z_1^1(\tau)^8 \right). \quad (7.37)$$

It is crucial that the fermions form groups of 16, so that the minus signs in Z_ψ^+ for 8 fermions, get squared into pluses, consistently with spacetime **Spin & Statistics**. This guarantees that the operators in $\mathfrak{A}_L$ have integral spin, rather than half-integral as required by modularity (Sect. 5.1). The full left-right fermionic partition function

$$Z_8(\tau)^2 Z_\psi^+(\tau)^*, \quad (7.38)$$

is then modular invariant, and we have a new consistent string theory.

Low-Lying Physical States As before, the lightest states on the right are the massless $\mathbf{8}_v \oplus \mathbf{8}_s$. On the left-moving side, in the NS-NS' sector the “normal order” constant sector, a , is given by the same computation as in (7.25), so $a = -1$. Thus the left-moving ground state is tachyonic, but it has no right-moving matching state, and so the physical spectrum is free of tachyons.

The first left-moving states surviving the GSO projection (7.35) are massless⁹

$$\begin{aligned} & \alpha_{-1}^i |0\rangle_{\text{NS,NS}'} \\ & \lambda_{-1/2}^A \lambda_{-1/2}^B |0\rangle_{\text{NS,NS}'} \quad 1 \leq A, B \leq 16 \text{ or } 17 \leq A, B \leq 32. \end{aligned} \quad (7.39)$$

There is a crucial difference with respect to the $SO(32)$ model: because there are separate GSO projections on each set of 16 fermions, A and B should come from the same set. The original $SO(32)$ symmetry is partly broken by the periodicity conditions (7.33), and we classify the states in representations of the preserved subgroup $SO(16) \times SO(16) \subset SO(32)$. The states in the second line of (7.39) transform in the (2-index antisymmetric) adjoint of the two $SO(16)$, each of dimension $16 \times 15/2 = \mathbf{120}$. In the left-moving R-NS' sector the shift a is

$$\text{R-NS}' : \quad \overbrace{\frac{8}{8}}^{h_S} - \frac{1}{24} \overbrace{\left(8 + \frac{32}{2} \right)}^{c_{Xi} + c_\lambda} \equiv \overbrace{0}^{h_S + h_c} \quad (7.40)$$

where now h_S is the weight of a $SO(16)$ spin field. Thus the R-NS' ground states are also massless, and transform in the spinor representation of the first $Spin(16)$. The spinor representation splits into two irreducible representations of definite chirality $\mathbf{256} = \mathbf{128}_s \oplus \mathbf{128}_c$. The GSO projection keeps the $\mathbf{128}_s$ only. The NS-R' sector produces a $\mathbf{128}_s$ of the *second* $SO(16)$, while the R-R' sector has no massless state.

⁹ The subscript NS, NS' means that both sets of left-moving fermions are in their NS-sector. Likewise the subscript R, NS' means that the first set of left-moving fermions is in the R-sector while the second set is in the NS-sector. Same story for NS, R' and R, R'.

Putting everything together, the $SO(8) \times SO(16) \times SO(16)$ representation content of the left-moving sector at the massless level is

$$(\mathbf{8}_v, \mathbf{1}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{120}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{1}, \mathbf{120}) \oplus (\mathbf{1}, \mathbf{128}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{1}, \mathbf{128}). \quad (7.41)$$

Combining these states with the right-moving $\mathbf{8}_v$ produces spacetime massless vectors in the $\mathbf{120} \oplus \mathbf{128}$ of each $SO(16)$. Consistency of the spacetime theory requires the massless vectors to transform in the *adjoint* of the gauge group. There is indeed a group—the exceptional Lie group E_8 [6]—which has a $SO(16)$ subgroup under which the E_8 adjoint representation $\mathbf{248}$ decomposes as $\mathbf{120} \oplus \mathbf{128}$. This second heterotic theory must have gauge group $E_8 \times E_8$ as we are going to demonstrate.

The $E_8 \times E_8$ Symmetry The world-sheet CFT has a full $E_8 \times E_8$ symmetry even though only an $SO(16) \times SO(16)$ subgroup is manifest in the fermionic description. The set of the $\mathbf{248}$ left-moving conserved currents of the (say) first E_8 is given by

$$\begin{aligned} \lambda^A \lambda^B(z) \quad A, B = 1, \dots, 16 \\ S_\alpha(z) \quad \text{the } \mathbf{128} \quad (-1)^{F_1} = +1 \text{ Spin}(16) \text{ spin operators.} \end{aligned} \quad (7.42)$$

Indeed, the Virasoro weight h_S of a $Spin(4n)$ spin operator is $n/4$ (Sect. 2.9) and hence the $Spin(16)$ spin operators $S_\alpha(z)$ have $(h, \bar{h}) = (1, 0)$, hence they are conserved currents ($\equiv$ holomorphic one-forms) whose contour integrals define the generators of a Lie algebra¹⁰ $\mathfrak{g}$ which must be *simple*¹¹ and hence should appear in Cartan's classification of finite-dimensional simple Lie algebras. Looking at Cartan's Table [7], we see that there is *precisely one* simple Lie algebra which decomposes as $\mathbf{120} \oplus \mathbf{128}$ under a $SO(16)$ subgroup, namely the exceptional Lie group E_8 .¹²

The above argument may be made explicit by looking to the OPEs between the left-moving currents (7.42). The two OPEs (2.491) and (2.493) with $d = 16$ yield

$$\lambda^A \lambda^B(z) S_\alpha(w) \sim \frac{1}{z-w} (\Gamma^{AB})_\alpha{}^\beta S_\beta(w) \quad (7.43)$$

$$S_\alpha(z) S_\beta(w) \sim \frac{C_{\alpha\beta}}{(z-w)^2} + \frac{1}{z-w} (\Gamma_{AB})_{\alpha\beta} \lambda^A \lambda^B(w), \quad (7.44)$$

where $\frac{1}{2}(\Gamma^{AB})_\alpha{}^\beta \equiv \frac{1}{2}(\Gamma^{[A}\Gamma^{B]})_\alpha{}^\beta$ are the $Spin(16)$ generators in the $\mathbf{128}$ representation. Together with the OPEs for the $SO(16)$ currents $\lambda^A \lambda^B$, the singular parts of the OPEs close on the currents (7.42) which then form a Kač-Moody algebra. The associativity of the OPE implies the Jacobi identity, and the zero modes of the

¹⁰ For convenience of the reader we review the well-known argument in BOX 7.1.

¹¹ **Proof:** Let $\mathfrak{j} \subset \mathfrak{g}$ be a proper Lie ideal. $[\lambda^A \lambda^B, \mathfrak{j}] \subset \mathfrak{j}$ implies that $\mathfrak{j}$ is a full $SO(16)$ representation. Neither the $\mathbf{120}$ nor the $\mathbf{128}$ are ideals, so $\mathfrak{j} = 0$ and the Lie algebra is simple.

¹² The adjoint representation of a classical Lie algebra of types $SU(n)$, $SO(n)$, $Sp(n)$ decomposes in classical representations of all subgroups i.e. the spin representation is never present. The $\mathbf{128}$ is a $SO(16)$ spin representation, so the full group should be exceptional, i.e. in the list G_2 , F_4 , E_6 , E_7 , E_8 . The full group contains $Spin(16)$, so its rank is at least 8 and the only possibility is E_8 .

currents generate a finite-dimensional Lie algebra which is a symmetry of the full CFT. By group theory this finite-dimensional Lie algebra must be E_8 : indeed, the standard presentation of the commutator relations which *define* the exceptional Lie algebra E_8 are constructed out of the $SO(16)$ gamma-matrices precisely as in Eqs. (7.43), (7.44),¹³ cf. [6, 9]. We shall return to this issue in Sect. 7.6. Here we limit to observe an elementary fact whose proof we leave to the reader:

Lemma 7.1 (Witt's theorem [10]) *Λ a positive-definite even lattice which is generated by the finite subset $R \subset \Lambda$ of its elements of square-length 2. Then R is the simply-laced reduced root system [11] of a simply-laced Lie algebra.*

Applying this result to the GSO projected lattice $(0) + (s)$ of $\mathfrak{spin}(16)$ (cf. Sect. 2.9) we get a simply-laced Lie algebra which turns out to be E_8 . This construction of the group E_8 in terms of $SO(16)$ current algebra is called the *direct construction* in the math literature, see Chaps. 6 and 7 of [6] or [9].

E_8 is a simply-laced Lie algebra of dimension **248** with Dynkin graph and Coxeter labels given by removing the unique extension node (the one with Coxeter label 1) from the rightmost graph in Fig. 2.3. Its Coxeter number $h \equiv h^\vee$ (equal to 1 plus the sum of the Coxeter labels) is then $h = 30$. The Sugawara central charge of the E_8 current algebra at level k is

$$c^{\text{sug}} = \frac{248k}{k+30} \equiv \frac{8(1+30)k}{k+30} \geq 8 \quad \text{with equality iff } k = 1. \quad (7.45)$$

Since the 16 left-moving MW fermions λ^A ($A = 1, \dots, 16$) have $c = 8$, we see that the matter left-moving CFT (a part for the embedding scalars X^μ) is just *two copies of the E_8 current algebra at level 1*. Put differently: after the GSO projection the bosonized $Spin(16)$ current algebra is nothing else than the Frenkel-Kač-Segal bosonization of the E_8 current algebra at level 1.

Summary of the Massless Sector The massless spectrum consists in the 10d $\mathcal{N} = 1$ supergravity multiplet plus an $E_8 \times E_8$ gauge supermultiplet. The $SO(8)$ spin times $E_8 \times E_8$ quantum numbers of the massless fields are

$$\begin{aligned} \text{gravity:} \quad & \overbrace{(\mathbf{1}, \mathbf{1}, \mathbf{1}) \oplus (\mathbf{28}, \mathbf{1}, \mathbf{1}) \oplus (\mathbf{35}, \mathbf{1}, \mathbf{1})}^{\text{bosons}} \oplus \overbrace{(\mathbf{56}_s, \mathbf{1}, \mathbf{1}) \oplus (\mathbf{8}_c, \mathbf{1}, \mathbf{1})}^{\text{fermions}} \\ \text{gauge:} \quad & \overbrace{(\mathbf{8}_v, \mathbf{248}, \mathbf{1}) \oplus (\mathbf{8}_v, \mathbf{1}, \mathbf{248})}^{\text{bosons}} \oplus \overbrace{(\mathbf{8}_s, \mathbf{248}, \mathbf{1}) \oplus (\mathbf{8}_s, \mathbf{1}, \mathbf{248})}^{\text{fermions}}. \end{aligned} \quad (7.46)$$

Note 7.1 We claim that the models with $G = SO(32)$ and $E_8 \times E_8$ are the *only* 10d SUSY heterotic theories. The algebra $\mathfrak{A}_L$ of the left-moving Fermi sector should be maximally local and spin-integral with $c = 16$. Let $\mathfrak{A}_L = \mathfrak{A}_1 \otimes \mathfrak{A}_2 \otimes \dots \otimes \mathfrak{A}_s$ where $\mathfrak{A}_k$ are maximally local spin-integral OPE algebras of irreducible, level-1,

¹³ This is obvious from 3d supergravity [8]: the maximal 3d SUGRA has target space $E_8/SO(16)$ and the fermions, which span the tangent space to the scalar manifold, should be in the spinor representation of the R-symmetry group $SO(16)$.

BOX 7.1 - All $(h, \tilde{h}) = (1, 0)$ Primaries Generate Global Symmetries

We recall why, in a unitary CFT, the set of $(h, \tilde{h}) = (1, 0)$ primary operators $\mathcal{O}^a(z)$ generate a Kac-Moody algebra while the associated charges

$$Q^a = \oint_C \frac{dz}{2\pi i} \mathcal{O}^a(z), \quad (\clubsuit)$$

generate a *reductive* (compact) finite-dimensional Lie algebra $\mathfrak{g}$. Let $|\mathcal{O}^a\rangle$ be the state corresponding to $\mathcal{O}^a$ which we assume, with no loss, to be Hermitian. One has

$$\|\tilde{L}_{-1}|\mathcal{O}^a\rangle\|^2 = \langle\mathcal{O}^a|[\tilde{L}_1, \tilde{L}_{-1}]|\mathcal{O}^a\rangle = 2\langle\mathcal{O}^a|\tilde{L}_0|\mathcal{O}^a\rangle = 0$$

since $\tilde{h} = 0$. Thus $\tilde{L}_{-1}|\mathcal{O}^a\rangle = 0$ or, in operators terms, $\bar{\partial}\mathcal{O}^a = 0$, so $\mathcal{O}^a(z)$ is a holomorphic (chiral left-moving) current. Saying that the charge Q^a in Eq. $(\clubsuit)$ is *conserved* means that the RHS depends on the contour C only through its homology class $[C]$ (in the world-sheet with operator insertion points deleted). Since $\mathcal{O}^a(z)$ is *holomorphic* in z , this follows at once from Cauchy residue theorem *as long as the integrand is well defined*, i.e. as long as $\mathcal{O}^a(z) dz$ is a global $(1,0)$ form on the world-sheet. This is equivalent to

$$\mathcal{O}^a(z')' = \mathcal{O}^a(z) (\partial z / \partial z'),$$

which is the definition of a primary with $h = 1$. For dimension reasons, the OPE has the form

$$\mathcal{O}^a(z) \mathcal{O}^b(w) \sim \frac{k^{ab}}{(z-w)^2} + \frac{f^{ab}{}_c \mathcal{O}^c(w)}{z-w},$$

for some tensors k^{ab} (symmetric) and $f^{ab}{}_c$ (antisymmetric). Integrating this relation we get

$$[Q^a, Q^b] = f^{ab}{}_c Q^c,$$

so the coefficients $f^{ab}{}_c$ satisfy the Jacobi identity, and hence are the structure constant of a finite-dimensional Lie algebra $\mathfrak{g}$. Associativity of the OPE implies that $k^{ab} \in \odot^2 \mathfrak{g}$ is a $\mathfrak{g}$ -invariant tensor. Unitarity implies that k^{ab} is positive-definite, so the Lie algebra $\mathfrak{g}$ is reductive (its exponential Lie group is compact), that is, $\mathfrak{g} = \mathfrak{a} \oplus \mathfrak{s}$ with $\mathfrak{a}$ Abelian and $\mathfrak{s}$ semisimple.

simply-laced current algebras with central charge c_k generated by $2c_k$ fermions out of the $32 \lambda^A$. We have $\sum_{k=1}^s c_k = 16$. By **Fact 5.2** each c_k is divisible by 8. Then either the current algebra is irreducible and $c_1 = 16$, or it has two irreducible factors and $c_1 = c_2 = 8$. These two solutions give $G = SO(32)$ and $E_8 \times E_8$ respectively. We shall give other two proofs of this claim in Sect. 7.6 and in Chap. 9.

7.3 Non-supersymmetric Heterotic Strings in 10d

All heterotic string theories can be constructed from a ‘basic’ model by the twisting procedure of Sect. 6.6.1: one gauges different discrete subgroups H of the symmetry G of the ‘basic’ theory. We work in two steps. First we consider twistings with no discrete torsion. For each subgroup $H \subset G$ we get a maximal local algebra $\mathfrak{A}_H$. The algebras $\mathfrak{A}_H$ which correspond to consistent models are the spin-integral ones. We classify the subgroups $H \subseteq G$ with this property, and list the corresponding consistent theories. In the second step we allow for discrete torsion: this yields one interesting model, the $SO(16) \times SO(16)$ heterotic string, which is tachyon-free [12].

The Diagonal Model

In any CFT the simplest modular invariant is the diagonal one. The fermionic factor of the partition function in the ‘basic’ heterotic model is this simple invariant:

$$\frac{1}{2} \left[Z_0^0(\tau)^{16} (Z_0^0(\tau)^*)^4 - Z_1^0(\tau)^{16} (Z_1^0(\tau)^*)^4 - \right. \\ \left. - Z_0^1(\tau)^{16} (Z_0^1(\tau)^*)^4 - Z_1^1(\tau)^{16} (Z_1^1(\tau)^*)^4 \right]. \quad (7.47)$$

The diagonal invariant corresponds to taking all 2d spinor fields (including $\tilde{\beta}$, $\tilde{\gamma}$) to have *identical* periodic conditions, i.e. to be sections of the same spin-structure line bundle on the torus. In the operator language: the world-sheet spinors are either all in the R sector or all in the NS sector, and we impose the diagonal GSO projection

$$(-1)^{F+\tilde{F}} = +1. \quad (7.48)$$

This theory is consistent except for the presence of a tachyon

$$\lambda^A e^{-\tilde{\phi}} |k\rangle, \quad m^2 \equiv -k^2 = -1, \quad (-1)^F = (-1)^{\tilde{F}} = -1, \quad (7.49)$$

which is a vector of $SO(32)$. The massless states are

$$\alpha_{-1}^i \tilde{\psi}_{-1/2}^j |e^{-\tilde{\phi}}, k\rangle, \quad \lambda_{-1/2}^A \lambda_{-1/2}^B \tilde{\psi}_{-1/2}^j |e^{-\tilde{\phi}}, k\rangle, \quad (7.50)$$

they are the graviton, dilaton, and 2-form field together with the $SO(32)$ gauge bosons. There are fermions in the spectrum but the lightest ones are at $m^2 = 2$

$$S_\sigma (\tilde{O} \tilde{S}_\alpha) e^{-\tilde{\phi}/2} |k\rangle, \quad k^2 = -2 \quad (7.51)$$

where S_σ is a left-moving $SO(32)$ spin field with $h = 32/16 = 2$, $\tilde{S}_\alpha$ is a right-moving $SO(9, 1)$ spin field, and $\tilde{O}$ is a right-moving NS superfield with $\tilde{h} = 1$.

The world-sheet QFT of this model has a symmetry $G \simeq \mathbb{Z}_2^{32} \subset O(32)$. The A -th generator of G , $(-1)^{F_A}$, flips the sign of the A -th left-moving fermion λ^A and leaves

invariant all others. Equation (7.48) implies that $(-1)^F \equiv \prod_A (-1)^{F_A}$ has the same effect as $(-1)^{\tilde{F}}$ which flips simultaneously the sign of all right-moving fermions.¹⁴

Twisting the Diagonal Model

We twist the diagonal model by a subgroup $H \subseteq G$. Consider a sector in which k out of the $32 \lambda^A$'s satisfy the R b.c. while the remaining $32 - k$ satisfy NS b.c. The left-moving ground state Virasoro weight h in such a sector is

$$h = \overbrace{\frac{k}{16}}^{\text{spin field}} - \overbrace{1}^{h_c}. \quad (7.52)$$

The oscillators raise h by multiples of $\frac{1}{2}$, so the h 's on the left-side are

$$h = \frac{k}{16} \mod \frac{1}{2}. \quad (7.53)$$

On the right-side we must take all fermions with the same boundary conditions to ensure BRST invariance, so the weights are multiples of $\frac{1}{2}$. Thus the condition $h - \tilde{h} \in \mathbb{Z}$ requires k to be a multiple of 8. Closure of the OPEs and spacetime **Spin & Statistics** require the stronger condition that k is a multiple of 16, cf. *Note 7.1*. We conclude that the operator algebra $\mathfrak{A}_H$, where $H \subset \mathbb{Z}_2^{32}$, is a maximal local algebra with $h - \tilde{h} \in \mathbb{Z}$ if and only if the following property holds:

Property (*) *the numbers of λ^A 's which commute and, respectively, anti-commute with the operator h are both multiples of 16 for all $h \in H$.*

This is to say that, in each h -twisted sector, the fermions with the same (periodic or anti-periodic) b.c. form groups of $16n$. *The model is SUSY if, in addition, $(-1)^F \in H$.*

SUSY Twistings

It follows from Sect. 7.2 that the twisted model is supersymmetric if and only if the projection $(-1)^{\tilde{F}} = 1$ is enforced, i.e. iff $(-1)^F \in H$. Consider twisting the diagonal theory by the group generated by $(-1)^F$. Together with the diagonal projection (7.48), it gives a total projection

$$\frac{1}{4}(1 + (-1)^{F+\tilde{F}})(1 + (-1)^F) = \frac{1}{4}(1 + (-1)^F)(1 + (-1)^{\tilde{F}}). \quad (7.54)$$

i.e. the projections $(-1)^F = (-1)^{\tilde{F}} = 1$ which define the supersymmetric $SO(32)$ heterotic string. If we consider twisting by groups generated by $(-1)^F$ and $(-1)^{F_1}$, where F_1 is the fermion number in the first block of 16 λ^A 's, we get the $E_8 \times E_8$ supersymmetric model.

¹⁴ Only the simultaneous flip of all right-moving fermions is consistent with BRST invariance.

Non-SUSY Twistings

Now we consider the non-SUSY twistings. If a subgroup $H \subset \mathbb{Z}_2^{32}$ has the property **(*)** but does not contain $(-1)^F$, each element $h \in H$, $h \neq 1$, anticommutes with precisely 16 left-moving fermions: we say that H has property **(**)**. We first construct a family of subgroups with property **(**)** and then show that all subgroups with this property are isomorphic to a group in the family.

We identify the index A of the fermion λ^A with a vector in the 5-dimensional vector space $\mathbb{F}_2^5$ over the field with two elements $\mathbb{F}_2$ [13]

$$\{1, \dots, 32\} \ni A \xrightarrow{f} \mathbf{d} \equiv (d_1, d_2, d_3, d_4, d_5) \in \mathbb{F}_2^5, \quad (7.55)$$

where the map f is *any* set-theoretic bijection between the two isomorphic sets. We define the operators $(-1)^{F_j}$ for $j = 1, \dots, 5$ to commute with the $\lambda^{\mathbf{d}}$ whose index $\mathbf{d} \in \mathbb{F}_2^5$ has j -th component $d_j = 0$, and anticommute with those with $d_j = 1$, that is,

$$(-1)^{F_j} \lambda^{\mathbf{d}} (-1)^{F_j} = (-1)^{d_j} \lambda^{\mathbf{d}}. \quad (7.56)$$

Let $V \in \mathbb{F}_2^5$ be a $\mathbb{F}_2$ -subspace, and consider the subgroup

$$H_V = \{(-1)^{\mathbf{n} \cdot \mathbf{F}} : \mathbf{n} \in V\} \subset \mathbb{Z}_2^{32}, \quad \mathbf{F} \equiv (F_1, F_2, F_3, F_4, F_5). \quad (7.57)$$

Lemma 7.2 *Let $h \in H_V$ be any element $h \neq 1$. Then h anticommutes (commutes) with precisely 16 out of the $32 \lambda^A$'s. That is, H_V has property **(**)** for all V .*

Proof We have $h = (-1)^{\mathbf{n} \cdot \mathbf{F}}$ for some $0 \neq \mathbf{n} \in V \subset \mathbb{F}_2^5$ and

$$h \lambda^{\mathbf{d}} h^{-1} = (-1)^{\mathbf{n} \cdot \mathbf{d}} \lambda^{\mathbf{d}} \quad (7.58)$$

So the indices $\mathbf{d}$ of the fermions commuting with h belong to the hyperplane $\mathbf{n} \cdot \mathbf{d} = 0$ in the 5-dimensional affine space $\mathbb{F}_2^5$ and hence have cardinality $|\mathbb{F}_2^4| = 16$. $\square$

Lemma 7.3 *Suppose $H \subset \mathbb{Z}_2^{32}$ has property **(**)**. Then $H \simeq H_V$ for some V .*

Proof We construct an index bijection $A \mapsto \mathbf{d}(A) \equiv (d_1(A), d_2(A), d_3(A), d_4(A), d_5(A))$ which does the job. Chose a set of generators $\{h_1, h_2, \dots, h_s\}$ of H and let $I_k \subset \{1, \dots, 32\}$ be the subset of indices such that $A \in I_k \Leftrightarrow h_k \lambda^A h_k^{-1} = -\lambda^A$. We claim that

$$|I_1 \cap I_2 \cap \dots \cap I_r| = 2^{5-r} \quad r = 1, 2, \dots, s. \quad (7.59)$$

The proof is by recursion in r . For $r = 1$ it is true. Replacing h_{r-1} with $h_{r-1} h_r$ as a generator of H , changes $I_{r-1} \rightsquigarrow I_{r-1} \Delta I_r$. Then, by the recursion hypothesis, we have

$$\begin{aligned} 2^{6-r} &= |I_1 \cap \dots \cap I_{r-2} \cap (I_{r-1} \Delta I_r)| = |I_1 \cap \dots \cap I_{r-2} \cap I_{r-1}| + |I_1 \cap \dots \cap I_{r-2} \cap I_r| - \\ &\quad - 2|I_1 \cap \dots \cap I_{r-1} \cap I_r| = 2^{6-r} + 2(2^{5-r} - |I_1 \cap \dots \cap I_{r-1} \cap I_r|) \end{aligned}$$

Equation (7.59) says that the map

$$d_k(A) = \begin{cases} 1 & \text{if } A \in I_k \\ 0 & \text{otherwise} \end{cases} \quad k = 1, \dots, 5. \quad (7.60)$$

is a bijection $\{1, \dots, 32\} \rightarrow \mathbb{F}_2^5$. The subspace $V \subset \mathbb{F}_2^{32}$ is the span of the first s basis vectors. $\square$

From the **Proof** we see that, modulo isomorphism, the subgroups H with property (**) depend only on the dimension s of the $\mathbb{F}_2$ -space V . Thus $s = 0, 1, 2, 3, 4, 5$ and we get six inequivalent non-SUSY models by twisting/projection with the operator

$$P_s \equiv \prod_{i=1}^s \frac{1 + (-1)^{F_i}}{2} \quad (7.61)$$

which defines a modular invariant, hence consistent, theory.

Spectrum and Gauge Group All six models so constructed have tachyons. Their vertices have the form $\lambda^A e^{ikX}$ for λ^A a fermion which is not projected out, i.e. which commutes with all $h \in H_V$. These are the ones whose index A satisfy $d_k(A) = 0$ for $k = 1, \dots, s$. We have precisely 2^{5-s} tachyons. The gauge group has the form $G = G_1 \times SO(2^{5-s})$, where $SO(2^{5-s})$ arises from the level-1 current algebra generated by the λ^A kept by the projection. The tachyons form a vector of $SO(2^{5-s})$ and are inert under G_1 .

By Eq. (7.52) and property (**) the ground states of all twisted sectors produce massless vectors by pairing with $\tilde{\psi}^\mu e^{-\tilde{\phi}}$ on the right. Before the projection they are Spin(16) spinors and have 2^8 components. Each of the s projections reduces them by half, so the number of massless vectors from twisted sectors are

$$\#_{\text{tw}} = (2^s - 1)(2^{8-s}). \quad (7.62)$$

The massless vector in the adjoint of G_1 , arising from the untwisted sector, have vertices $\lambda^A \lambda^B \tilde{\psi}_\mu e^{-\tilde{\phi}} e^{ikX}$ which are kept by the projection while λ^A is not. This means $\mathbf{d}(A), \mathbf{d}(B)$ have the same first s components, which are not all zero. Their number is

$$\#_{\text{untw}} = \frac{1}{2}(2^s - 1)2^{5-s}(2^{5-s} - 1) \quad (7.63)$$

So that

$$\dim G_1 \equiv \#_{\text{untw}} + \#_{\text{tw}} = 2^{4-s}(2^s - 1)(2^{5-s} + 15). \quad (7.64)$$

We also know that for $s < 5$ the total rank of G is 16, so that $\text{rank } G_1 = 16 - 2^{4-s}$. Indeed the number of λ^A which have the same periodicity in all sectors is even, and we may pair them into complex fermions which we bosonize as in Sect. 2.9.2. Cartan algebra currents of G are $\partial\phi_a$ where ϕ_a ($a = 1, \dots, 16$) are the chiral scalars which bosonize the λ^A . The $s = 5$ case can be formulated only in terms of 2d MW fermions,

Table 7.2 Non-SUSY twistings of the diagonal model (*no torsion*)

$s \equiv \dim V$	$\dim G_1$	$\text{rank } G_1$	$G \equiv G_1 \times SO(2^{5-s})$
0	0	0	$SO(32)$
1	248	8	$E_8 \times SO(16)$
2	276	12	$SO(24) \times SO(8)$
3	266	14	$E_7 \times E_7 \times SO(4)$
4	255	15	$SU(16) \times SO(2)$
5	248	8	E_8

and has no simple bosonization. We get Table 7.2. We do not comment further on these models which, having tachyons, are not very interesting.

Exercise 7.1 Work out the details of the massless spectrum of the $s = 5$ model.

Note 7.2 One may wonder about twisting by a discrete subgroup $H \subset O(32)$ with $H \not\subset \mathbb{Z}_2^{32}$; no such consistent subgroup is known.

It remains another possibility, namely introduce a non-trivial discrete torsion in the H -projection along the lines of Sect. 6.6.1.

Adding Discrete Torsion: the $SO(16) \times SO(16)$ Heterotic String

Recall from Sect. 6.6.1 that when we twist a string model (more generally a CFT), specified by a maximal local algebra $\mathfrak{A}$, by an Abelian symmetry group H , producing a maximal local algebra $\mathfrak{A}_H$ which satisfies the condition $h - \tilde{h} \in \mathbb{Z}$, we can construct other such algebras $\mathfrak{A}_{H,\varepsilon}$ by introducing a *discrete torsion* given by a bilinear map

$$\varepsilon: \wedge^2 H \rightarrow U(1). \quad (7.65)$$

$\mathfrak{A}_{H,\varepsilon}$ is also a maximal local algebra with $h - \tilde{h} \in \mathbb{Z}$, i.e. a consistent string theory.

We generalize the discussion of the previous paragraph by allowing non-trivial discrete torsion in the H -twisting where $H \subset \mathbb{Z}_2^{32}$ is a subgroup satisfying (*). Note that, since H is a 2-torsion group,¹⁵ the phase $\varepsilon(h_1, h_2)$ is actually a *sign*.

When H satisfies the stronger condition (**)—that is, if $H \simeq H_{\mathbb{F}_2^s}$ for some s —we do not get anything *essentially* new. In facts, let us look at the resulting massless spectrum. Since $\varepsilon(1, h) = 1$ for all h , in the untwisted sector the projection remains the same as before, and so does the untwisted sector spectrum. We still get 2^{5-s} tachyons in the vector representation of the gauge group $SO(2^{5-s})$, and the same untwisted vectors as before. Before the projection, the ground states of the h -twisted sector, still form a spinor of $Spin(16)$, to which we apply the projector

$$P_h = \prod_{i=1}^s \frac{1 + \varepsilon(h, (-1)^{F_i}) (-1)^{F_i}}{2} \quad (7.66)$$

¹⁵ I.e. a group such that $h^2 = 1$ for all $h \in H$.

Each of the s projection reduces the number of states by two, and this is independent of the signs $\varepsilon(h, (-1)^{F_i})$ induced by the discrete torsion. Therefore we get exactly the same massless spectrum as before.

To get something interesting we need an H which satisfy $(*)$ with $(-1)^F \in H$. There are only two of them. The group generated by $(-1)^F$ which yields the $SO(32)$ SUSY model, and the group generated by $(-1)^F$ and $(-1)^{F_1}$ which gives the $E_8 \times E_8$ SUSY theory. Since $\varepsilon(h_1, h_2)$ is alternating, it is trivial unless H contains at least *two* independent generators. Thus we remain with only one interesting candidate, the group H producing the $E_8 \times E_8$ model. In this situation there is a unique alternating ε , up to equivalence. We identify $H \simeq \mathbb{F}_2^2$ by writing $h \in H$ in the form

$$h = \exp[\pi i(k_2 F_1 + l_2 \tilde{F})], \quad (k, l) \in \mathbb{F}_2^2. \quad (7.67)$$

Then $\varepsilon: \wedge^2 \mathbb{F}_2^2 \rightarrow \mathbb{F}_2 \simeq \{\pm 1\}$ is just the canonical symplectic form in the $\mathbb{F}_2$ -plane

$$\varepsilon(h_1, h_2) = (-1)^{k_1 l_2 - k_2 l_1} \equiv (-1)^{\det(k_i, l_i)} \quad (7.68)$$

which is manifestly bilinear and alternating. For the twisting group $H \equiv \langle (-1)^F, (-1)^{F_1} \rangle$ the fermionic periodic conditions are labelled by three elements of $\mathbb{Z}_2$, $(\alpha_1, \alpha'_1, \tilde{\alpha})$

$$\lambda^A(w + 2\pi) = -(-1)^{\alpha_1} \lambda^A(w), \quad A = 1, \dots, 16 \quad (7.69)$$

$$\lambda^A(w + 2\pi) = -(-1)^{\alpha'_1} \lambda^A(w), \quad A = 17, \dots, 32 \quad (7.70)$$

$$\tilde{\psi}^\mu(w + 2\pi) = -(-1)^{\tilde{\alpha}} \tilde{\psi}^\mu(w), \quad (7.71)$$

or, writing collectively the fermionic fields as Ψ ,

$$\Psi(w + 2\pi) = -(-1)^{\alpha'_1} e^{\pi i[(\alpha_1 - \alpha'_1)F_1 + (\tilde{\alpha} - \alpha'_1)\tilde{F}]} \Psi(w) e^{\pi i[(\alpha_1 - \alpha'_1)F_1 + (\tilde{\alpha} - \alpha'_1)\tilde{F}]}, \quad (7.72)$$

that is,

$$h_2 = \exp[\pi i(\alpha_1 - \alpha'_1)F_1 + \pi i(\tilde{\alpha} - \alpha'_1)\tilde{F}]. \quad (7.73)$$

Comparing with Eqs. (7.67), (7.68), we see that inserting in the path integral the discrete torsion (7.68) we get the following sector-wise redefinition of the projectors

$$e^{i\pi \tilde{F}} \rightarrow e^{i\pi \tilde{F}} (-1)^{\alpha_1 - \alpha'_1} \quad (7.74)$$

$$e^{i\pi F_1} \rightarrow e^{i\pi F_1} (-1)^{\alpha'_1 - \tilde{\alpha}} \quad (7.75)$$

$$e^{i\pi F'_1} \rightarrow e^{i\pi F'_1} (-1)^{\alpha'_1 - \tilde{\alpha}} (-1)^{\alpha_1 - \alpha'_1} \equiv e^{i\pi F'_1} (-1)^{\alpha_1 - \tilde{\alpha}}. \quad (7.76)$$

By construction the OPEs are consistent. The sectors which are kept are (the 3 entries in each parenthesis correspond, respectively, to $\tilde{\psi}^\mu$, the first 16 λ^A , the last 16 λ^A)

Table 7.3 Massless spectrum and physical vertices of the $SO(16) \times SO(16)$ heterotic string

Sector	Vertices	$SO(8)_{\text{spin}} \times SO(16) \times SO(16)$ reprs.
(NS+, NS+, NS+)	$\partial X^\mu \tilde{\psi}^v e^{-\tilde{\phi}} e^{ikX}$	$(\mathbf{1}, \mathbf{1}, \mathbf{1}) \oplus (\mathbf{28}, \mathbf{1}, \mathbf{1}) \oplus (\mathbf{35}, \mathbf{1}, \mathbf{1})$
	$\chi^a \chi^b \tilde{\psi}^v e^{-\tilde{\phi}} e^{ikX}$	$(\mathbf{8}_v, \mathbf{120}, \mathbf{1})$
	$\eta^a \eta^b \tilde{\psi}^v e^{-\tilde{\phi}} e^{ikX}$	$(\mathbf{8}_v, \mathbf{1}, \mathbf{120})$
(R+, NS−, NS−)	$\chi^a \eta^b \tilde{S}_\alpha e^{-\tilde{\phi}/2} e^{ikX}$	$(\mathbf{8}_s, \mathbf{16}, \mathbf{16})$
(R−, R−, NS+)	$S_\sigma \tilde{S}_\alpha e^{-\tilde{\phi}/2} e^{ikX}$	$(\mathbf{8}_c, \mathbf{128}_c, \mathbf{1})$
(R−, NS+, R−)	$S'_\sigma \tilde{S}_\alpha e^{-\tilde{\phi}/2}$	$(\mathbf{8}_c, \mathbf{1}, \mathbf{128}_c)$

$$\begin{array}{cccc}
(\text{NS}+, \text{NS}+, \text{NS}+) & (\text{NS}-, \text{NS}-, \text{R}+) & (\text{NS}-, \text{R}+, \text{NS}-) & (\text{NS}+, \text{R}-, \text{R}-) \\
(\text{R}+, \text{NS}-, \text{NS}-) & (\text{R}-, \text{NS}+, \text{R}-) & (\text{R}-, \text{R}-, \text{NS}+) & (\text{R}+, \text{R}+, \text{R}+)
\end{array} \quad (7.77)$$

From Eq. (7.77) we easily extract the main physical features of this model:

- The only sectors which contain gravitini are $(\text{R}\pm, \text{NS}+, \text{NS}+)$. Since these sectors are projected out, the theory has no gravitini and hence no spacetime SUSY;
- Tachyons are projected out, since they belong to sectors $(\text{NS}-, \text{NS}\pm, \text{NS}\mp)$;
- Massless vectors are in sectors $(\text{NS}+, \text{NS}+, *)$, $(\text{NS}+, *, \text{NS}+)$, $(\text{NS}+, \text{NS}-, \text{NS}-)$. The only such sector which survives the projection is $(\text{NS}+, \text{NS}+, \text{NS}+)$ which contains massless vectors in the adjoint of $SO(16) \times SO(16)$. *The space-time gauge group is $SO(16) \times SO(16)$.*

The $SO(16) \times SO(16)$ heterotic model is an example of a tachyon-free theory which is non-supersymmetric. The massless spectrum is described in Table 7.3. For clarity in the table we wrote the left-moving fermions of the two $SO(16)$ current algebras as χ^a, η^a , respectively with $a = 1, \dots, 16$ instead of λ^A ; S_σ and S'_σ are the spin-fields of the two $SO(16)$ current algebras of negative chirality.

7.4 Heterotic Strings: The Bosonic Construction

We know from our study of winding states in the bosonic string (cf. Chap. 6) and bosonization of free chiral CFTs (Sect. 2.5) that we may consider *independent* left- and right-moving 2d chiral scalar fields. Then we may construct a heterotic string by combining 26 left-moving scalars with 10 right-moving ones together with their right-moving fermionic super-partners $\tilde{\psi}^\mu$. These world-sheet d.o.f. produce the correct Virasoro central charges as they are literally the fusion of the left-moving sector of the bosonic string in 26 dimensions with the right-sector of a 10d superstring. However, for such a world-sheet theory to be a *bona fide* CFT, we need to ensure that the resulting left-right *asymmetric* 2d theory is modular invariant, or, equivalently, that its operator algebra $\mathfrak{A}$ is maximally local and spin-integral. This requires a generalization

of the projections we performed in Sect. 2.9.2 in the context of bosonization of level-1 $SO(8n)$ current algebras.

The main issue is the allowed spectrum of the bosonic zero-modes (i.e. left/right momenta) $k_{L,R}$; as always we use their dimensionless version

$$l_{L,R} \stackrel{\text{def}}{=} \sqrt{\alpha'/2} k_{L,R}. \quad (7.78)$$

Recall that an ordinary *non-compact* scalar corresponds to a pair of left- and right-moving chiral scalars with momenta

$$l_L^\mu = l_R^\mu = l^\mu \quad (7.79)$$

which take *continuous* values. We assume our heterotic string to move in $2 \leq d \leq 10$ non-compact dimensions. We remain with $26 - d$ left-moving chiral scalars and $10 - d$ right-moving ones whose momenta

$$l \equiv (l_L^m, l_R^n), \quad d \leq m \leq 25, \quad d \leq n \leq 9, \quad (7.80)$$

take value in some “lattice”¹⁶

$$\Gamma \subset \mathbb{R}^{36-2d}. \quad (7.81)$$

In this book by a *lattice* of rank n , $\Gamma \subset \mathbb{R}^n$, we always mean a discrete subgroup of the Abelian group $\mathbb{R}^n$, such that $\Gamma \otimes_{\mathbb{Z}} \mathbb{R} \simeq \mathbb{R}^n$, which in addition is endowed with an integral, non-degenerate, symmetric, bilinear form $\circ: \Gamma \times \Gamma \rightarrow \mathbb{Z}$ (the “inner product”). The *signature* (r, s) of a lattice is the signature of the corresponding inner product extended to $\mathbb{R}^n$. One has $r + s \equiv n$. Positive-definite lattices, of signature $(n, 0)$, are called *Euclidean lattices*.

In the heterotic context the bilinear form $\circ$ is defined by the chiral scalars’ OPEs. Consider the vertex

$$V_l(z, \bar{z}) = (\text{cocycle}) e^{il_L \cdot X_L + il_R \cdot X_R} \quad (7.82)$$

which generates the state of the chiral scalars’ CFT with momentum $l \in \Gamma$ and oscillator numbers $N = \tilde{N} = 0$. From the OPE (6.66) we have

$$V_l(e^{2\pi i} z) V_{l'}(0) = e^{2\pi i l \circ l'} V_l(z) V_{l'}(0) \quad (7.83)$$

where $\circ$ is the indefinite inner product

$$l \circ l' \stackrel{\text{def}}{=} l_L \cdot l'_L - l_R \cdot l'_R, \quad (7.84)$$

¹⁶ By “lattice” *between quotes* we mean a discrete subgroup of the additive group $\mathbb{R}^n$ which generates $\mathbb{R}^n$ as a vector space. A lattice *without quotes* carries additional structures as explained in the main text. That Γ is an Abelian group follows from closure of the OPE; discreteness follows from locality for 2d scalars which *do not* represent ordinary non-compact dimensions.

of signature $(26 - d, 10 - d)$. The bilinear form $\circ$ controls the mutual locality between two such vertices: the operator algebra of the chiral scalars is local iff the bilinear form $\circ$ is *integral* in Γ

$$\circ: \Gamma \times \Gamma \rightarrow \mathbb{Z}, \quad (7.85)$$

i.e. iff the pair $(\Gamma, \circ)$ is a *lattice* according to our definition.

The heterotic string mass formulae in d dimensions take the form

$$\text{on the left} \quad \frac{\alpha'}{4} m^2 = \frac{1}{2} l_L^2 + N - 1 \quad (7.86)$$

$$\text{on the right} \quad \frac{\alpha'}{4} m^2 = \frac{1}{2} l_R^2 + \begin{cases} \tilde{N} - \frac{1}{2} & \text{NS sector} \\ \tilde{N} & \text{R sector} \end{cases} \quad (7.87)$$

Modular Invariance I

To get a consistent model the vertex algebra $\mathfrak{A}$ should be maximal local of integral spin. To simplify matters we construct only *supersymmetric* heterotic models, which are the ones of main physical interest. As reviewed in Sect. 7.2, this requires the usual GSO projection $\exp[\pi i \tilde{F}_{\text{GSO}}] = +1$ on the right. With this projection we have

$$\mathfrak{A} = \mathfrak{A}_{\text{nc}} \otimes \mathfrak{A}_{\tilde{\psi}} \otimes \mathfrak{A}_{\Gamma} \quad (7.88)$$

where $\mathfrak{A}_{\text{nc}}$ is the CFT operator algebra of the non-compact scalars X^μ and $b, \tilde{b}, c, \tilde{c}$ ghosts, $\mathfrak{A}_{\tilde{\psi}}$ is the CFT algebra of the fermions $\tilde{\psi}^\mu, \tilde{\psi}^i$ and $\tilde{\beta}, \tilde{\gamma}$ ghosts, and $\mathfrak{A}_{\Gamma}$ the CFT algebra of the $(26 - d, 10 - d)$ left-right chiral scalars with momenta in the lattice Γ . By effect of the (ordinary) GSO projection both algebras $\mathfrak{A}_{\text{nc}}, \mathfrak{A}_{\tilde{\psi}}$ are maximal local of integral spin. Therefore $\mathfrak{A}$ is a maximal local spin-integral algebra—and hence the associated heterotic model is modular invariant—iff the chiral scalars' algebra $\mathfrak{A}_{\Gamma}$ is maximal local.

By definition $\mathfrak{A}_{\Gamma}$ is generated by the operators V_l in (7.82) with $l \equiv (l_L, l_R) \in \Gamma$. An operators V_l with $l \in \mathbb{R}^{36-2d}$ is mutually local to *all* operators in $\mathfrak{A}_{\Gamma}$ iff

$$l \in \{l \in \mathbb{R}^{36-2d} : l \circ k \in \mathbb{Z} \ \forall k \in \Gamma\} \equiv \Gamma^\vee, \quad (7.89)$$

so $\mathfrak{A}_{\Gamma}$ is maximal local if and only if the lattice Γ is *self-dual* (see BOX 2.10)

$$\Gamma^\vee = \Gamma. \quad (7.90)$$

The conformal spin of V_l is

$$h(V_j) - \tilde{h}(V_l) = \frac{1}{2}(l_L^2 - l_R^2) \equiv \frac{1}{2}l \circ l \quad (7.91)$$

and $\mathfrak{A}_\Gamma$ has integral spin iff $l \circ l \in 2\mathbb{Z}$ for all l , i.e. if Γ is an *even* lattice (cf. **BOX 2.10**). We conclude that our SUSY heterotic model is modular invariant iff Γ is a *self-dual even lattice*. This is the same conclusion we got in Sect. 6.5 when studying Narain compactification¹⁷ of the bosonic string. The only difference is that in Sect. 6.5 the signature of Γ was (k, k) while now has the form (r, s) with $r - s = 16$. The proof of maximal locality with integral spin is independent of the signature, and works identically in the two cases.

Exercise 7.2 Extend the analysis to cover non-SUSY heterotic models.

Modular Invariance II

We arrive at the same conclusion looking to the explicit partition function on the torus. The first factor in Eq. (7.88) has a modular invariant partition function: the non-compact scalars lead to the modular invariant partition function (cf. Eqs. (4.55)–(4.58) or Eq. (5.99))

$$Z_X(\tau, \bar{\tau})^{d-2}. \quad (7.92)$$

while the partition function of the right-moving fermions is (cf. Chap. 5)

$$(Z_\psi^+(\tau))^* \quad \text{with modular transformations} \quad \begin{cases} (Z_\psi^+(\tau+1))^* = e^{-2\pi i/3} (Z_\psi^+(\tau))^* \\ (Z_\psi^+(-1/\tau))^* = (Z_\psi^+(\tau))^* \end{cases} \quad (7.93)$$

The partition function of the $(26-d, 10-d)$ left/right scalars with momenta in Γ is

$$Z(\tau, \bar{\tau})_\Gamma = \frac{1}{\eta(\tau)^{26-d} \bar{\eta}(\bar{\tau})^{10-d}} \sum_{(l_L, l_R) \in \Gamma} q^{l_L^2/2} \bar{q}^{l_R^2/2}. \quad (7.94)$$

Under $T: \tau \rightarrow \tau + 1$ the pre-factor transforms as

$$\begin{aligned} \frac{1}{\eta(\tau)^{26-d} \bar{\eta}(\bar{\tau})^{10-d}} &\rightarrow \frac{e^{-2\pi i(26-d)/24 + 2\pi i(10-d)/24}}{\eta(\tau)^{26-d} \bar{\eta}(\bar{\tau})^{10-d}} \equiv \\ &\equiv e^{2\pi i/3} \frac{1}{\eta(\tau)^{26-d} \bar{\eta}(\bar{\tau})^{10-d}}, \end{aligned} \quad (7.95)$$

while the terms of the sum over Γ in (7.94) transform as

$$q^{l_L^2/2} \bar{q}^{l_R^2/2} \rightarrow (-1)^{l \circ l} q^{l_L^2/2} \bar{q}^{l_R^2/2} \quad (7.96)$$

so that the sum over momenta is invariant precisely iff $l \circ l \in 2\mathbb{Z}$ for all $l \in \Gamma$, i.e. iff Γ is an *even lattice*. Thus for Γ even

$$Z(\tau+1, \bar{\tau}+1)_\Gamma = e^{2\pi i/3} Z(\tau, \bar{\tau})_\Gamma, \quad (7.97)$$

¹⁷ Narain introduced his compactifications in the present heterotic context [14, 15].

in agreement with Eq. (5.36). A computation identical to the one in Eq. (6.181) gives

$$Z(-1/\tau, -1/\bar{\tau})_{\Gamma} = \frac{1}{\text{Vol}(\Gamma)} Z(\tau, \bar{\tau})_{\Gamma^{\vee}}. \quad (7.98)$$

so S -invariance requires Γ to be self-dual, in agreement with (7.90). We conclude:

The SUSY heterotic partition function

$$\left(Z_X(\tau, \bar{\tau}) \right)^{d-2} Z(\tau, \bar{\tau})_{\Gamma} \left(Z_{\psi}^+(\tau) \right)^*$$

is modular invariant if and only if the lattice Γ , with its inner product $\circ$ of signature $(26 - d, 10 - d)$, is even and self-dual, that is:

- (i) $l \circ l \in 2\mathbb{Z}$ for all $l \in \Gamma$,
- (ii) $\Gamma = \Gamma^{\vee} \stackrel{\text{def}}{=} \left\{ x \in \Gamma \otimes_{\mathbb{Z}} \mathbb{Q} \mid x \circ y \in \mathbb{Z} \ \forall y \in \Gamma \right\}$

Thus all even self-dual lattice Γ of signature $(26 - d, 10 - d)$ yields a supersymmetric heterotic string moving in d -dimensional Minkowski space, invariant under a SUSY extension of the Poincaré group with 16 supercharges.¹⁸ To build all such theories we have to construct all even self-dual lattices with the appropriate signature. The basic result in the classification of even self-dual lattices is **Theorem 6.1**: lattices with the required properties *exist for all d* since $(26 - d) - (10 - d) = 0 \bmod 8$.

Lattices Versus Matrices Choosing a system of generators $\{e_{\alpha} \in \mathbb{R}^n\}_{\alpha=1}^n$ for $\Gamma \subset \mathbb{R}^n$ (a.k.a. an *integral basis*) we get an isomorphism (of free Abelian groups)

$$\Gamma \equiv \bigoplus_{\alpha=1}^n \mathbb{Z} e_{\alpha} \simeq \mathbb{Z}^n, \quad (7.99)$$

and we can rewrite the inner product $\circ$ as the bilinear form

$$\circ: \mathbb{Z}^n \times \mathbb{Z}^n \rightarrow \mathbb{Z}, \quad \circ: (x, y) \mapsto x^t C y, \quad (7.100)$$

in terms of its matrix $C_{\alpha\beta}$ (with respect to the chosen integral basis)

$$C_{\alpha\beta} \stackrel{\text{def}}{=} e_{\alpha} \circ e_{\beta}, \quad \alpha, \beta = 1, 2, \dots, n, \quad (7.101)$$

¹⁸ The right-moving world-sheet currents which produce the space-time SUSY Ward identities (cf. Sect. 3.8) transform in the $\mathbf{16}_s$ of $Spin(9, 1)$. These currents are unaffected by the compactification, so in d dimensions still have 16 supercharges which organize themselves in spinors of $Spin(d - 1, 1)$.

which is a symmetric $n \times n$ matrix with integral entries. The lattice Γ is *even* iff the diagonal entries $C_{\alpha\alpha}$ of its matrix are even integers, and it is *self-dual* iff the determinant of $C_{\alpha\beta}$ is a unit of $\mathbb{Z}$, i.e. iff

$$\det C = \pm 1. \quad (7.102)$$

The signature (r, s) is specified by the number r (resp. s) of positive (resp. negative) eigenvalues of the matrix $C \equiv (C_{\alpha\beta})$.

We identify the lattice Γ with the $\mathbb{Z}$ -equivalence class of the matrix C . Two integral matrices C and C' belong to the same $\mathbb{Z}$ -equivalence class iff $C' = ACA^t$ for some $A \in GL(n, \mathbb{Z})$. The $\mathbb{Z}$ -equivalence class of C is independent of the choice of integral basis for Γ , and all matrices in the class represent $\circ$ in some integral basis.

7.5 Classification of Even Self-dual Lattices

This section is mathematical, the reader may prefer to skip it in a first reading.

We say that two lattices $\Gamma, \Gamma' \subset \mathbb{R}^n$ of signature (r, s) ($n \equiv r + s$) are *equivalent*, written $\Gamma' \sim \Gamma$, if they are related by an overall $O(r, s; \mathbb{R})$ rotation, i.e. if, given integral bases $\{e_\alpha\}$ of Γ , there exists an element $O \in O(r, s; \mathbb{R})$ such that the vectors

$$e'_\alpha = O \cdot e_\alpha, \quad \alpha = 1, 2, \dots, n, \quad (7.103)$$

form an integral basis of Γ' . We already know from Sect. 6.5 that equivalence in this mathematical sense does *not* imply physical equivalence of the corresponding string models. As explained there, physical equivalence is a *finer* relation which is preserved only by rotations (7.103) with O in the maximal compact subgroup

$$O \in O(r) \times O(s) \subset O(r, s; \mathbb{R}). \quad (7.104)$$

These aspects being clarified, we state a more precise version of **Theorem 6.1**:

Theorem 7.1 (E.g. [10, 16]) *The following facts hold:*

- (1) *Even self-dual lattices exist in signature (r, s) if and only if $r - s = 0 \bmod 8$;*
- (2) *If the necessary condition (1) is satisfied, and $rs \neq 0$ (i.e. the lattice is indefinite), there is precisely one equivalence class of even self-dual lattices;*
- (3) *In the definite case, i.e. in signatures $(8k, 0)$ and $(0, 8k)$, there are only finitely many even self-dual lattices of the given rank $8k$, up to equivalence;*

- (4) In the definite case, let $\text{Aut}(\Gamma)$ be the automorphism group of the definite, even, self-dual lattice Γ .¹⁹ We write Ω_m for the set of inequivalent, even, self-dual lattices of signature $(2m, 0)$. The Minkowski-Siegel mass formula holds:

$$\sum_{\Gamma \in \Omega_m} \frac{1}{|\text{Aut}(\Gamma)|} = \frac{|B_m|}{2m} \prod_{\ell=1}^{m-1} \frac{|B_{2\ell}|}{4\ell} \quad (7.105)$$

where B_j is the j -th Bernoulli number (see **BOX 7.2**).

Since $|\text{Aut}(\Gamma)| \geq 2$, the number $\mathfrak{N}_{8k}$ of inequivalent definite, even, self-dual lattices in $\mathbb{R}^{8k}$ is at least $\mathfrak{N}_{8k} > 2 f(k)$, where the function $f(k)$ is defined in **BOX 7.2**. We stress that this is a *very poor* lower bound on the number of lattices, as the table in **BOX 7.2** illustrates. The inequivalent definite, even, self-dual lattices in $\mathbb{R}^{32}$ are already more than 80 millions, and their number grows with the rank n at least as rapidly as

$$2 f(n/8) \approx n^{n^2} \cdot \exp[O(n^2)] \quad \text{as } n \equiv 8k \rightarrow \infty. \quad (7.106)$$

The E_8 Cartan Matrix Let C be the Cartan matrix of a simply-laced, semi-simple Lie algebra $\mathfrak{g}$ of rank r [7, 11, 18]. By definition, C is a symmetric, positive-definite,²⁰ integral, $r \times r$ matrix with

$$\det C = |Z_{\mathfrak{g}}|, \quad (7.107)$$

where $|Z_{\mathfrak{g}}|$ is the order of the center $Z_{\mathfrak{g}} \subset G$ of the simply-connected compact Lie group G whose Lie algebra is $\mathfrak{g}$. One has

$$|Z_{A_n}| = n + 1, \quad |Z_{D_n}| = 4, \quad |Z_{E_r}| = 9 - r, \quad r = 6, 7, 8, \quad (7.108)$$

so that the Cartan matrix is unimodular, $\det C = 1$, if and only if

$$\mathfrak{g} = \overbrace{E_8 \oplus E_8 \oplus \cdots \oplus E_8}^{n/8 \text{ copies}}, \quad (7.109)$$

expressing the fact that E_8 is the only simply-laced, simply-connected, simple Lie group which has trivial center, i.e. such that its fundamental representation is the adjoint one.²¹ The rank of a Cartan matrix C which yields a positive-definite, even, self-dual lattice is a multiple of 8, in agreement with **Theorem 7.1**.

¹⁹ The group $\text{Aut}(\Gamma)$ is necessarily *finite*, being both discrete and compact. $\text{Aut}(\Gamma)$ is compact, because it is a *closed* subgroup of the compact group $O(8k)$, and it is *discrete* because it is a subgroup of the discrete group $GL(n, \mathbb{Z})$.

²⁰ For the positive-definiteness of C see **Appendix 2** of Chap. 2.

²¹ By a *fundamental* representation of a *compact semi-simple* Lie group G we mean a smallest faithful representation F . Then all finite-dimensional representations V (in particular all the unitary irreducible ones) are sub-representations of some tensor power of F , i.e. $V \subset F^{\otimes \ell}$ for some ℓ .

BOX 7.2 - Bernoulli Numbers, ζ -Function, and All That

The Bernoulli numbers are *rational*s which are best defined in terms of their generating function

$$\sum_{k=0}^{\infty} B_k \frac{t^k}{k!} = \frac{t}{e^t - 1}$$

from which it follows that $B_{2s+1} = 0$ for $s > 0$. The non-zero Bernoulli numbers alternate in sign. Several explicit formulae for B_n are known, see e.g. Sect. 24.6 of [17]; for instance

$$B_n = \sum_{k=0}^n \frac{1}{k+1} \sum_{j=0}^k (-1)^j \binom{k}{j} j^n.$$

Let $\zeta(s)$ be the Riemann ζ -function. For m a positive integer we have

$$\zeta(2m) = \frac{(2\pi)^{2m}}{(2m)!} |B_{2m}|, \quad \zeta(-m) = -\frac{B_{m+1}}{m+1}$$

Since $\zeta(s) \rightarrow 1$ as $s \rightarrow +\infty$ along the real axis, one has the asymptotic estimate

$$\frac{|B_{2m}|}{4m} \approx \frac{(2m-1)!}{2(2\pi)^{2m}} \quad \text{as } m \rightarrow \infty \quad (*)$$

Consider the function

$$f(k) \stackrel{\text{def}}{=} \frac{|B_{4k}|}{8k} \prod_{\ell=1}^{4k-1} \frac{|B_{2\ell}|}{4\ell},$$

Equation (*) and the Stirling formula yield the following asymptotic expression of $f(k)$ for large k

$$\log f(k) \approx 2 \log G(8k) + \text{sub-leading}$$

where $G(s)$ is Barnes' double-gamma function [17] defined by the functional equation

$$G(s+1) = \Gamma(s) G(s) \quad \text{with } G(1) = 1.$$

As $s \rightarrow \infty$ one has

$$\log G(s) \approx \frac{1}{2} s^2 \log s + \text{sub-leading}$$

For illustration, we present a list of *approximate* values of $f(k)$ for the first few k 's

k	1	2	3	4
$f(k)$	1.435×10^{-9}	2.489×10^{-18}	7.937×10^{-15}	4.031×10^7

We write E_8 for the even, self-dual Euclidean lattice of rank 8 of matrix²²

$$E_8 \stackrel{\text{def}}{=} \begin{bmatrix} 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 \end{bmatrix} \quad (7.110)$$

Together with Witt's **Lemma 7.1** this implies that all positive, even, self-dual lattices which are not direct sums of E_8 lattices cannot be generated by vectors of squared-length 2 only. **Theorem 7.1(2)** gives

Corollary 7.1 Γ an indefinite, even, self-dual lattice of signature (r, s) . Then

$$\Gamma \sim \frac{r-s}{8} E_8 \oplus \min\{r, s\} H \quad (7.111)$$

where H (as in Sect. 6.5) is the even, self-dual lattice of signature $(1, 1)$ of matrix

$$H = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad (7.112)$$

and the short-hand notation qC (with $q \in \mathbb{Z}$) stands for the direct sum $C^{\oplus|q|}$ (resp. $(-C)^{\oplus|q|}$) of $|q|$ copies of the matrix C (resp. $-C$) for $q > 0$ (resp. $q < 0$).

Even Self-Dual Lattices of Signature $(8, 0)$

The lower rank in which there is a positive-definite, even, self-dual lattice is 8. Such a lattice is given by a symmetric, integral 8×8 matrix C with 2's on the main diagonal and determinant 1. We already know one example: the E_8 lattice (7.110).

Proposition 7.1 E_8 is the only even, self-dual Euclidean lattice of rank 8.

Proof We show that the contribution of the E_8 lattice to the LHS of the Minkowski-Siegel mass formula (7.105) saturates its RHS. Of course, $|\text{Aut}(E_8)| \equiv |\text{Weyl}(E_8)|$, so the **Proposition** claims that the following numerical identity holds

$$\frac{1}{|\text{Weyl}(E_8)|} \stackrel{?}{=} \frac{|B_4|}{8} \prod_{j=1}^3 \frac{|B_{2j}|}{4j} \equiv \frac{|B_2| \cdot |B_4|^2 \cdot |B_6|}{2^{10} \cdot 3} \equiv \frac{1}{2^{14} \cdot 3^5 \cdot 5^2 \cdot 7}, \quad (7.113)$$

²² Cf. the last diagram in Fig. 2.3 with the extension node ($\equiv$ the rightmost one) removed: $C_{\alpha\beta} = 2\delta_{\alpha\beta} - I_{\alpha\beta}$ where $I_{\alpha\beta}$ is the incidence matrix of the resulting Dynkin graph.

BOX 7.3 - The Order of the Weyl Group

We give two formulas for the order $|\text{Weyl}(\mathfrak{g})|$ of the Weyl group of a simple Lie algebra $\mathfrak{g}$ of rank r . G stands for the simply-connected Lie group with Lie algebra $\mathfrak{g}$ and Γ_{root} for its root lattice.

Lemma 7.4 (COROLLAIRE 1, Chap. V, Sect. 6 No. 2 of [11]) *Let $d_i = \ell_i + 1$ ($i = 1, \dots, r$) be the degrees of the Casimir invariants of $\mathfrak{g}$ (equivalently, the ℓ_i 's are the exponents of $\mathfrak{g}$). Then*

$$|\text{Weyl}(\mathfrak{g})| = \prod_{i=1}^r d_i$$

Lemma 7.5 (PROPOSITION 7, Chap. VI, Sect. 2 No. 4 of [11]) *Let m_i the Dynkin labels of $\mathfrak{g}$ (see BOX 2.8 and Figs. 2.2, 2.3a, 2.3b, 2.3c), and*

$$z \equiv [\Gamma_{\text{root}}^\vee : \Gamma_{\text{root}}] \equiv |Z(G)|$$

Then

$$|\text{Weyl}(\mathfrak{g})| = r! z \prod_{i=1}^r m_i.$$

Examples. Using the second expression

$$|\text{Weyl}(\mathfrak{so}(2r))| = r! \cdot 4 \cdot 1^3 \cdot 2^{r-3} \equiv 2^{r-1} r!$$

$$|\text{Weyl}(\mathfrak{so}(32))| = 2^{30} \cdot 3^6 \cdot 5^3 \cdot 7^2 \cdot 11 \cdot 13 = 685\,597\,979\,049\,984\,000$$

$$|\text{Weyl}(E_8)| = 8! \cdot 1 \cdot (1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6) \cdot (4 \cdot 2) \cdot 3 \equiv 2^{14} \cdot 3^5 \cdot 5^2 \cdot 7 = 696\,729\,600$$

where we used (see e.g. Sect. 24.2(iv) of [17])

$$|B_2| = \frac{1}{2 \cdot 3}, \quad |B_4| = \frac{1}{2 \cdot 3 \cdot 5}, \quad |B_6| = \frac{1}{2 \cdot 3 \cdot 7}. \quad (7.114)$$

Equality (7.113) follows from standard formulae for the order of Weyl group: see BOX 7.3. $\square$

Note 7.3 We already constructed the positive, even, self-dual lattice of rank 8 when studying the level-1 $Spin(16)$ current algebra in Sect. 2.9.2. There we had 16 left-moving MW fermions, λ^A ($A = 1, \dots, 16$), bosonized their current algebra with 8 left-moving chiral scalars $\phi(z)_i$, and GSO-projected the Hilbert space to the $(-1)^F = +1$ sector which keeps the cosets (o) and (s) . The momenta of the chiral scalars take value in a GSO lattice $\Gamma_{\text{GSO}\pm}$ ($\Gamma_{\mathfrak{so}(16)} \subset \Gamma_{\text{GSO}\pm} \subset \Gamma_{\mathfrak{so}(16)}^\vee$) which is even and self-dual. This lattice was identified with the E_8 root lattice around **Lemma 7.1**.

Proposition 7.1 yields a deeper rationale for this identification.

Even Self-Dual Lattices of Signature (16, 0)

We already know *two* inequivalent even, self-dual lattices, namely $E_8 \oplus E_8$ and $Spin(32)^+$, the lattice obtained by the GSO projection of the $Spin(32)$ weight lattice which keeps the two cosets (o) and (s) (notations as in Sect. 2.9.2). We shall revisit these two positive-definite lattices in the next section.

Proposition 7.2 *$E_8 \oplus E_8$ and $Spin(32)^+$ are the only positive-definite, even, self-dual lattices of rank 16, up to equivalence.*

Proof We check that the contributions to the LHS of the Minkowski-Siegel formula from these two lattices saturate its RHS. One has

$$|\text{Aut}(E_8 \oplus E_8)| = 2 |\text{Weyl}(E_8)|^2, \quad |\text{Aut}(Spin(32)^+)| = |\text{Weyl}(SO(32))|, \quad (7.115)$$

so the contribution to the LHS of (7.105) from these two lattices is²³

$$\begin{aligned} \frac{1}{2^{29} \cdot 3^{10} \cdot 5^4 \cdot 7^2} + \frac{1}{2^{30} \cdot 3^6 \cdot 5^3 \cdot 7^2 \cdot 11 \cdot 13} &= \\ &= \frac{2 \cdot 11 \cdot 13 + 3^4 \cdot 5}{2^{30} \cdot 3^{10} \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13} = \frac{691}{2^{30} \cdot 3^{10} \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13}, \end{aligned} \quad (7.116)$$

while the RHS of (7.105) is

$$\frac{|B_2| \cdot |B_4| \cdot |B_6| \cdot |B_8|^2 \cdot |B_{10}| \cdot |B_{12}| \cdot |B_{14}|}{2^{22} \cdot 3^2 \cdot 5 \cdot 7} \quad (7.117)$$

Using Eq. (7.114) and

$$|B_8| = \frac{1}{2 \cdot 3 \cdot 5}, \quad |B_{10}| = \frac{5}{2 \cdot 3 \cdot 11}, \quad |B_{12}| = \frac{691}{2 \cdot 3 \cdot 5 \cdot 7 \cdot 13}, \quad |B_{14}| = \frac{7}{2 \cdot 3} \quad (7.118)$$

we see that the two rational numbers (7.116) and (7.117) indeed agree. $\square$

7.6 SUSY Heterotic Strings in $d = 10$ (Bosonic Form)

We consider first the maximal number of non-compact dimensions, $d = 10$. In this situation the product $\circ$ is positive-definite and Γ is an *even self-dual Euclidean lattice* of rank 16. As reviewed in the previous section, even self-dual Euclidean lattices exist only if the rank is a multiple of 8 [10, 16], and in rank 16 there are precisely two inequivalent such lattices (**Proposition 7.2**). We adopt the physical notations and write them Γ_{16} and $\Gamma_8 \times \Gamma_8$, respectively (in the math notation of Sect. 7.5 they are $Spin(32)^+$ and $E_8 \oplus E_8$, respectively). The lattice Γ_{16} is (cf. Sect. 2.9.2)

$$\Gamma_{16} \stackrel{\text{def}}{=} \left\{ (n_1, \dots, n_{16}) \in \left(\frac{1}{2}\mathbb{Z}\right)^{16} \left| \begin{array}{l} n_i = n_j \pmod{1}, \\ \sum_i n_i = 0 \pmod{2} \end{array} \right. \right\} \quad (7.119)$$

²³ Note that 691 is a prime integer.

The lattice Γ_8 is defined analogously,

$$\Gamma_8 \stackrel{\text{def}}{=} \left\{ (n_1, \dots, n_8) \in \left(\frac{1}{2}\mathbb{Z}\right)^8 \left| n_i = n_j \bmod 1, \sum_i n_i = 0 \bmod 2 \right. \right\}. \quad (7.120)$$

Comparing with the discussion of the $SO(8n)$ current algebra in Sect. 2.9.2, we see that Γ_8 (resp. Γ_{16}) is nothing else than the GSO projected lattice

$$\Gamma_{\text{GSO}} \equiv (o) + (s) \subset \Gamma_{\text{weight}} \quad (7.121)$$

for $SO(16)$, resp. $SO(32)$. (o) was defined to be the $SO(8n)$ root lattice $\Gamma_{\mathfrak{so}(8n)}$ while (s) was the coset

$$(s) = \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \dots, \frac{1}{2}\right) + \Gamma_{\mathfrak{so}(8n)}. \quad (7.122)$$

We know from Sect. 2.9.2 that the lattice Γ_{GSO} is *self-dual Euclidean* for all $SO(8n)$. When n is even Γ_{GSO} is also *even*: indeed for $SO(8n)$ one has

$$(o) \cdot (o) \in 2\mathbb{Z}, \quad (s) \cdot (s) \equiv \left(\frac{1}{2}, \dots, \frac{1}{2}\right) \cdot \left(\frac{1}{2}, \dots, \frac{1}{2}\right) \bmod 2 \equiv n \bmod 2. \quad (7.123)$$

In the heterotic string, as in the bosonic string, the left-moving zero-point energy is -1 , equal to the weight of the standard b, c Fermi sea. Hence

$$\frac{\alpha'}{4} m^2 = \frac{1}{2} l_L^2 + N - 1, \quad (7.124)$$

and the left-moving part of the *massless* vertices have one of the three forms²⁴

$$\partial X^\mu \quad \text{or} \quad \partial X^m \quad \text{or} \quad e^{il_L \cdot X} \quad \text{with } l_L^2 = 2. \quad (7.125)$$

Tensoring with the usual right-moving massless vertices in the $\mathbf{8}_v \oplus \mathbf{8}_s$ of the transverse Lorentz group $SO(8)$, ∂X^μ gives the 10d $\mathcal{N} = 1$ supergravity multiplet (graviton, dilaton, 2-form field B , MW gravitino, MW dilatino of opposite chirality), while

$$\partial X^m \quad \text{as well as} \quad e^{il_L \cdot X} \quad \text{with } l_L^m \in \Gamma \text{ and } l_L^2 = 2, \quad (7.126)$$

produce massless vectors and gluinos which should form precisely *one copy* of the adjoint representation of the spacetime gauge group G .

The $16 \partial X^m$ form a maximal set of commuting world-sheet currents corresponding to the ‘KK’ momenta in the compactified directions. In view of the relation between 2d chiral $(1, 0)$ currents and spacetime gauge symmetries (see **BOX 7.1**), we have to identify the ∂X^m with the chiral currents in the Cartan subalgebra $\mathfrak{h}$ of the gauge group G , which then has rank 16. The momenta l_L^m in the massless vertices (7.126) are the

²⁴ For the rest of the section we set $\alpha' = 2$ to simplify notations.

$\mathfrak{h}$ -charges of the gauge vectors $\notin \mathfrak{h}$, that is, the operators $e^{il_L \cdot X(z)}$, of weights $(1, 0)$, are by definition *the chiral currents associated to the root generators $\{E_\alpha\}_{\alpha \in \Delta(\mathfrak{g})}$ of the gauge Lie algebra*²⁵

$$\mathfrak{g} = \mathfrak{h} \oplus \left(\bigoplus_{\alpha \in \Delta(\mathfrak{g})} \mathbb{R} E_\alpha \right). \quad (7.127)$$

The non-Cartan chiral $(1,0)$ currents $e^{il_L \cdot X(z)}$ are in one-to-one correspondence with the points of length-square 2 in the lattice Γ , cf. Eq. (7.126). That is,

When $d = 10$ the root system R of the spacetime gauge group G (of rank 16) is the set of elements in Γ of squared-length 2

The spacetime gauge group G is realized on the world-sheet as a *level-1* current algebra with Virasoro central charge equal to its rank, $c = r(\mathfrak{g}) \equiv 16$, so G must be simply-laced by Eq. (2.387). In particular all roots have the same squared-length 2.

Points of Squared-Length 2 in Γ_{16} For the lattice $\Gamma_{16} = (o) \oplus (s)$ we have

$$l \in (s) \Rightarrow l \cdot l \geq 16 \times \frac{1}{4} \equiv 4, \quad (7.128)$$

so $l^2 = 2$ implies $l \in (o)$ (the $SO(32)$ root lattice) and l^m is a $SO(32)$ root. Hence Γ_{16} leads to the gauge Lie algebra $\mathfrak{g} = \mathfrak{so}(32)$. We have recovered from the bosonic construction the $SO(32)$ SUSY heterotic string we already constructed using 2d free fermions in Sect. 7.2. This is hardly a surprise since the lattice Γ_{16} arises from the bosonization of 32 free Majorana fermions and then taking the GSO projection, cf. Sect. 2.9.2, which is precisely what we did in Sect. 7.2.

Points of Squared-Length 2 in Γ_8 Now we have

$$l \in (s) \Rightarrow l \cdot l \geq 8 \times \frac{1}{4} \equiv 2, \quad (7.129)$$

with equality iff $l^m = \pm \frac{1}{2}$ for all m . Since $\sum_m l^m \in 2\mathbb{Z}$, the roots belonging to the coset $(s) \equiv \Gamma_8 / \Gamma_{\mathfrak{so}(16)}$ transform in the $\mathbf{128}_s$ of $\mathfrak{so}(16)$. Adding the adjoint of $\mathfrak{so}(16)$ (from ∂X^m and $(o) \equiv \Gamma_{\mathfrak{so}(16)}$), the gauge vectors associated to one Γ_8 factor transform in the $\mathbf{120} \oplus \mathbf{128}_s \equiv \mathbf{248}$, which we know to be the adjoint of E_8 . The even self-dual lattice $\Gamma_8 \times \Gamma_8$ produces the gauge group $E_8 \times E_8$ and we recover the $E_8 \times E_8$ heterotic string. Indeed, the bosonic construction is just the Frenkel-Kač-Segal bosonization of the $E_8 \times E_8$ KM algebra at level $k = 1$. This result is pretty obvious from the viewpoint of classification theory of even, self-dual, Euclidean lattices of ranks 8 and 16 discussed in the previous section.

²⁵ $\Delta(\mathfrak{g})$ stands for the root system of the Lie algebra $\mathfrak{g}$ [7, 11, 18].

7.7 Toroidal Compactifications

Now we consider SUSY heterotic models moving in $d < 10$ non-compact dimensions. All even self-dual lattices Γ of indefinite signature (r, s) are equivalent; therefore, for $d < 10$ all even self-dual lattices of signature $(26 - d, 10 - d)$ can be obtained from any given one, Γ_0 , by an ambient $O(26 - d, 10 - d; \mathbb{R})$ rotation. We may take Γ_0 to be the direct sum of one of the 10d lattices (Γ_{16} or $\Gamma_8 \times \Gamma_8$) and $(10 - d)H$ where H is the rank-2 lattice (7.112). The standard choice is

$$\Gamma_0 = (10 - d) H \oplus E_8 \oplus E_8, \quad (7.130)$$

cf. (7.111). Γ_0 describes the 10d $E_8 \times E_8$ heterotic string compactified on $T^{10-d} \equiv (S^1)^{10-d}$ with all radii set at the self-dual value $R = \sqrt{\alpha'}$. Any lattice Γ of the form

$$\Gamma = O\Gamma_0, \quad O \in O(26 - d, 10 - d, \mathbb{R}) \quad (7.131)$$

defines a consistent heterotic string compactification on T^{10-d} . As in the bosonic string, Sect. 6.5, we have the *physical* equivalence

$$O_1 O O_2 \Gamma_0 \sim O\Gamma_0, \quad (7.132)$$

where

$$O_1 \in O(26 - d, \mathbb{R}) \times O(10 - d, \mathbb{R}), \quad O_2 \in O(26 - d, 10 - d, \mathbb{Z}). \quad (7.133)$$

Moduli Space: Geometry

The moduli space $\mathcal{M}_d$ of toroidal compactifications of the supersymmetric heterotic string to d dimensions is the double coset

$$\left(O(26 - d) \times O(10 - d) \right) \backslash O(26 - d, 10 - d, \mathbb{R}) / O(26 - d, 10 - d, \mathbb{Z}) \quad (7.134)$$

that is, the quotient of the connected and simply-connected Riemannian symmetric space of non-compact type²⁶

$$\widetilde{\mathcal{M}}_d = \left(O(26 - d) \times O(10 - d) \right) \backslash O(26 - d, 10 - d, \mathbb{R}) \quad (7.135)$$

by the arithmetic subgroup²⁷ $O(26 - d, 10 - d, \mathbb{Z})$ of its isometry group. This is precisely the kind of space which naturally appears as scalars' target space in a supergravity model with more than 8 supercharges [8]. The $\mathbb{R}$ -rank of the symmetric space is [19]

²⁶ Hence the *covering moduli space* $\widetilde{\mathcal{M}}_d$ is a Hadamard space diffeomorphic to $\mathbb{R}^{\dim \mathcal{M}_d}$.

²⁷ The precise definition is $O(26 - d, 10 - d, \mathbb{Z}) \stackrel{\text{def}}{=} \{g \in GL(36 - 2d, \mathbb{Z}) : g^t \eta g = \eta\}$ with η the diagonal matrix with $(26 - d)$ 1's and $(10 - d)$ -1's along the diagonal.

$$\mathbb{R}\text{-rank} = \min\{26 - d, 10 - d\} = 10 - d. \quad (7.136)$$

The $\mathbb{Q}$ -rank of the corresponding $\mathbb{Q}$ -algebraic group²⁸ $O(26 - d, 10 - d, \mathbb{Q})$ coincides with its $\mathbb{R}$ -rank, hence is maximal. In particular, for all $d \leq 9$ $\mathcal{M}_d$ is *non-compact of finite volume* (by Godement's criterion [24]), while for $d \leq 8$ $\mathcal{M}_d$ is *Mostow rigid* and *Margulis super-rigid* [24]. These deep Number-Theoretic properties are related to the swampland conjectures, see [8] Chap. 4.

***T*-Duality Group**

As in bosonic case, the arithmetic subgroup $O(26 - d, 10 - d, \mathbb{Z})$ is called the *the *T*-duality group* and, in our conventions, it acts on the right.²⁹ For $d \leq 8$ the *T*-duality group is essentially³⁰ determined by the requirement that the moduli space has finite volume which (by definition [24]) means that the *T*-duality group is a *lattice* in the Lie group $O(26 - d, 10 - d, \mathbb{R})$ (cf. [8] Chap. 4).

Just as in the bosonic string, the *T*-duality group is Borel dense in the real Lie group $O(26 - d, 10 - d, \mathbb{R})$ (cf. **Theorem 6.3**), and the two groups cannot be distinguished by *any* order parameter transforming in a finite-dimensional representation.

Gauge Symmetry

To avoid any confusion, we write X^μ for the non-compact scalars, and Y^m for the left/right compact chiral ones whose zero-modes are valued in the lattice Γ .

We look for the unbroken space-time gauge symmetry. First of all, there are $26 - d$ massless gauge bosons with vertices of the form

$$\partial Y^m \tilde{\psi}^\mu e^{-\tilde{\phi}} e^{ik \cdot X}, \quad k^2 = 0 \quad (7.137)$$

and $10 - d$ ones with vertices of the form

$$\partial X^\mu \tilde{\psi}^\mu e^{-\tilde{\phi}} e^{ik \cdot X}, \quad k^2 = 0. \quad (7.138)$$

These vectors correspond to the 16 symmetries in the Cartan subalgebra of the 10d gauge Lie algebra together with $10 - d$ Kaluza Klein gauge bosons, $G_{\mu m}$, and the $10 - d$ vectors from the mixed components of the 2-form, $B_{\mu m}$. In addition, we have a massless gauge boson with vertex

$$e^{il_L \cdot Y_L} \tilde{\psi}^\mu e^{-\tilde{\phi}} e^{ik \cdot X}, \quad k^2 = 0 \quad (7.139)$$

for every point in the lattice Γ such that

$$l_L^2 = 2, \quad l_R = 0. \quad (7.140)$$

²⁸ For algebraic groups over a field see e.g. [20–22]. For their arithmetic subgroups see [21, 23–25]. For $\mathbb{Q}$ -rank see [24].

²⁹ This is the opposite of the convention mostly used in the math literature.

³⁰ That is, up to commensurability [24].

Note that there are no gauge bosons with $l_R \neq 0$ since the mass of such a state would be at least $\frac{1}{2}l_R^2$ due to the right-moving GSO projection.

For *generic* boosts O in Eq. (7.131), corresponding to *generic points* in the covering moduli $\widetilde{\mathcal{M}}_d$, there are *no* elements in Γ with $l_R = 0$, and hence no additional gauge bosons. The *generic* gauge group is simply

$$U(1)^{36-2d}. \quad (7.141)$$

At special points in $\widetilde{\mathcal{M}}_d$ the space-time gauge symmetry gets *enhanced* to a non-Abelian group. Obvious allowed gauge groups are

$$SO(32) \times U(1)^{2(10-d)} \quad \text{or} \quad E_8 \times E_8 \times U(1)^{2(10-d)}, \quad (7.142)$$

obtained by compactifying the corresponding 10d heterotic theory on a generic torus T^{10-d} with no Wilson line: (7.142) is the same gauge group we would get by trivial dimensional reduction in field theory. However, just as in the bosonic string, there are also special points in $\widetilde{\mathcal{M}}_d$ where we get *stringy* enhanced gauge symmetries. For instance, the indefinite, even, self-dual lattice

$$\Gamma_* \equiv \left\{ (m_i) \in \left(\frac{1}{2}\mathbb{Z} \right)^{26-d, 10-d} \mid m_i = m_j \bmod 1, \sum_j m_j = 0 \bmod 2 \right\}, \quad (7.143)$$

defined in analogy with Γ_{16} , Γ_8 , has a gauge symmetry maximally enhanced to

$$SO(52-2d) \times U(1)^{10-d}. \quad (7.144)$$

All gauge groups obtained in this way have rank $36-2d$. This is the maximum rank we may get in string perturbation theory, but we shall see that non-perturbative effects may lead to even larger gauge groups. The general gauge group has the form

$$G = G_L \times U(1)^{10-d} \quad \text{rank } G_L = 16-d, \quad (7.145)$$

where the $U(1)^{10-d}$ gauge vectors have vertices (7.138).

As in the bosonic string, the low energy physics near a point of enhanced symmetry is described by the Higgs mechanism. Globally on $\widetilde{\mathcal{M}}_d$ the story is a bit subtler since, say, the group $G_L = E_8 \times E_8$ cannot arise by standard symmetry breaking from the ‘largest’ left group $G_L^{\text{lar}} = SO(52-2d)$ with adjoint Higgs fields. The subtlety is related to the arithmetic nature of the Higgs phenomenon in the gravity context.

★ Arithmetics of Gauge Symmetry Enhancement

We saw in Sect. 6.4 that for a single non-chiral compact scalar the point p_{sd} in the *covering* moduli space $\widetilde{\mathcal{M}} \equiv O(1, 1, \mathbb{R})/O(1)^2 \simeq \mathbb{R}_{\geq 0}$ which corresponds to self-dual radius $R = \sqrt{\alpha'}$ was precisely the unique point fixed by non-trivial elements $h \in O(1, 1, \mathbb{Z})$. The isotropy subgroup $\text{Fix}(p_{\text{sd}}) \subset O(1, 1, \mathbb{Z})$ was identified with the Weyl group of the enhanced non-Abelian gauge

symmetry $SU(2)_L \times SU(2)_R$ at the special point $p_{\text{sd}} \in \widetilde{\mathcal{M}}$. This observation taught us that the T -duality group of the bosonic Narain compactifications is a *gauge symmetry* of the full string theory.

These observations remain true in the heterotic contexts for the left-moving side (which is the bosonic string). The story goes as follows. The covering moduli space $\widetilde{\mathcal{M}}_d$, Eq. (7.135), is a Hadamard space, hence by Cartan theorem [26], each finite subgroup of $O(26-d, 10-d; \mathbb{Z})$ has a non-empty fixed set in $\widetilde{\mathcal{M}}_d$, which is a totally geodesic submanifold when its dimension is positive.³¹ The ‘special points’ in the moduli spaces $\mathcal{M}_d$ are the images of the fixed loci of some finite subgroup of the T -duality group. These are the points where the generic gauge group $U(1)^{36-2d}$ enhances to a non-Abelian gauge group. Again we expect a relation between the isotropy subgroup of the enhancement point and the Weyl group of the enhanced gauge group at that point.

Let $p \in \mathcal{M}_d$ be a point fixed by a non-trivial subgroup $\mathbf{Fix}(p) \subset O(26-d, 10-d; \mathbb{Z})$. -1 acts trivially, so the group of actual interest is $\mathbf{Fix}(p)/\{\pm 1\}$. Let $O_p \in O(26-d, 10-d; \mathbb{R})$ be a representative of the point p ; concretely this means that

$$\gamma O_p = O_p(u_\gamma \times v_\gamma), \quad u_\gamma \times v_\gamma \in SO(26-d) \times O(10-d) \quad \text{for all } \gamma \in \mathbf{Fix}(p). \quad (7.146)$$

We thus get a group homomorphism $\pi_L: \mathbf{Fix}(p)/\{\pm 1\} \rightarrow O(26-d)/\{\pm 1\}$, given by $\gamma \mapsto \pm u_\gamma$, which we call the *left-projection of the T -isotropy group*. Since

$$\mathbf{Fix}(p) \subset O(26-d, 10-d; \mathbb{Z}) \subset O(26-d, 10-d; \mathbb{Q}), \quad (7.147)$$

the homomorphism is defined over $\mathbb{Q}$, so its image is a finite subgroup of $O(26-d; \mathbb{Q})$. All finite subgroups of $O(26-d; \mathbb{Q})$ are conjugate³² to subgroups of $O(26-d; \mathbb{Z})_Q \subset GL(26-d; \mathbb{Z})$ which preserve a positive-definite quadratic form $Q: \mathbb{Z}^{26-d} \rightarrow \mathbb{Z}$. Therefore the image of $\mathbf{Fix}(p)$ (resp. $\mathbf{Fix}(p)/\{\pm 1\}$) is isomorphic to a subgroup of $O(26-d; \mathbb{Z})_Q$ (resp. $O(26-d; \mathbb{Z})_Q/\{\pm 1\}$). The left-projection of the T -isotropy group $W \subset O(26-d; \mathbb{Z})_Q/\{\pm 1\}$ is a *normal subgroup*.³³ This normal subgroup is the candidate *Weyl group for the enhanced non-Abelian gauge group* at $p \in \widetilde{\mathcal{M}}$. Hence all allowed gauge symmetries in heterotic toroidal compactifications have Weyl groups $W \triangleleft O(26-d; \mathbb{Z})_Q/\{\pm 1\}$ for some Euclidean lattice Q of rank $(26-d)$.

The simplest possibility is that Q is the ‘trivial’ lattice I_{26-d} , i.e. $Q(x_i) = \sum_i x_i^2$, so that $O(26-d; \mathbb{Z})_Q \equiv O(26-d; \mathbb{Z})$. We note the following isomorphisms³⁴

$$O(26-d; \mathbb{Z}) \simeq G(2, 1, 26-d) \simeq \text{Weyl}(\mathfrak{so}(53-2d)) \quad (7.148)$$

$$O(26-d; \mathbb{Z})/\{\pm 1\} \simeq G(2, 2, 26-d) \simeq \text{Weyl}(\mathfrak{so}(52-2d)) \quad (7.149)$$

This result has a clear interpretation. The lattice Γ_* corresponds to a gauge group G_L whose Weyl group is the full group $O(26-d; \mathbb{Z})/\{\pm 1\}$. The corresponding points $p_* \in \widetilde{\mathcal{M}}_d$ are the ones with $\pi_L \mathbf{Fix}(p_*)/\{\pm 1\} \simeq \text{Weyl}(\mathfrak{so}(52-2d))$. It is convenient to take this lattice as the reference one Γ_0 ; with this new convention $p_* \in \widetilde{\mathcal{M}}_d$ is the class of $1 \in O(26-d, 10-d; \mathbb{R})$.

Now consider a subgroup $W \triangleleft O(26-d; \mathbb{Z})/\{\pm 1\}$. Let $\mathcal{P}_W \subset \widetilde{\mathcal{M}}_d$ be the locus of points with isotropy $\mathbf{Fix}(p)/\{\pm 1\} \supset W$. We have $p_* \in \mathcal{P}_W$ for all such W . Thus to study the gauge enhancement patterns with Weyl group a normal subgroup $W \triangleleft O(26-d; \mathbb{Z})/\{\pm 1\}$ it suffices to work in a small neighborhood of 1 , that is, at the linearized level around p_* , working with scalars which live in the tangent space to $\widetilde{\mathcal{M}}_d$ at the identity, which is isomorphic to the -1 eigenspace of the Cartan

³¹ Conversely an infinite subgroup may fix only ‘cusps’ at infinite distance in the moduli space. This aspect of Cartan’s theorem is related to the *distance conjecture* of the swampland program.

³² Cf. Sect. 1.3.1 of [27].

³³ $\mathbf{Fix}(p)$ is normal as a subgroup of $O(26-d, 10-d; \mathbb{R})$, hence normal in any subgroup containing $\mathbf{Fix}(p)$ such as the $\mathbb{Q}$ -conjugate of $O(26-d; \mathbb{Q}) \subset O(26-d; \mathbb{Q}) \times O(10-d; \mathbb{Q}) \subset O(26-d, 10-d; \mathbb{R})$ which contains it.

³⁴ $G(p, q, n)$ (with $q|p$) are the Shephard-Todd imprimitive reflection groups [28–34].

involution acting on the Lie algebra $\mathfrak{so}(26-d, 10-d)$ [19, 26, 35] which makes $10-d$ copies of the Cartan algebra $\mathfrak{h}$ of G_* (rotated by the ‘R-symmetry’ $O(10-d)$). Clearly they correspond to the v.e.v. $\langle \Phi_i \rangle$ of $(10-d)$ effective Higgs fields in the adjoint of $SO(52-2d)$ which commute in a static configuration (so have a leading potential $\propto \text{tr}([\Phi_i, \Phi_j])^2$ as in Eq. (6.255)). Thus, at first order in a neighborhood of $p_* \in \tilde{\mathcal{M}}_d$ the pattern of gauge symmetry enhancement may be understood in conventional field-theoretical terms as the Higgs breaking of the parent $SO(52-2d)$ gauge symmetry by Higgs fields in the adjoint.

However there is a lot more than that. Here the *arithmetic* nature of the Higgs phenomenon in quantum gravity plays the crucial role. The lattice Q needs *not* to be $\mathbb{Z}$ -equivalent³⁵ to the trivial lattice I_{26-d} . Other allowed lattices give rise to different maximal gauge groups which are rank $(26-d)$ Lie groups which cannot arise from the adjoint Higgs breaking of $SO(52-2d)$. For instance, we may have $Q = I_{10-d} \oplus E_8 \oplus E_8$, which leads to $G_L = SO(10-d) \times E_8 \times E_8$ which cannot be obtained from $SO(52-2d)$ by the conventional field theoretic Higgs mechanism, but do arise in the arithmetic Higgs set-up.

Moduli Space $\mathcal{M}_d$: Target-Space Interpretation

The dimension of the moduli space (7.134) is

$$\dim \mathcal{M}_d = (26-d)(10-d). \quad (7.150)$$

As in the bosonic string, this result has a simple interpretation. The moduli space—or, rather, its symmetric cover $\tilde{\mathcal{M}}_d$, Eq. (7.135)—parametrizes the background fields of the 10d theory consistent with d -dimensional super-Poincaré symmetry with 16 supercharges. Then the T -duality group $O(26-d, 10-d, \mathbb{Z})$ is a gauge symmetry which identifies all backgrounds in the same orbit. The components of the metric and antisymmetric tensor in the compact factor T^{10-d} of space-time, G_{mn} and B_{mn} , give $(10-d)^2$ moduli. In addition, there are Wilson lines, i.e. constant backgrounds for the gauge fields A_m . Due to the potential term $\text{Tr}([A_m, A_n]^2)$ in a vacuum configuration the A_m fields in different directions commute, and hence the A_m ’s may be simultaneously rotated in the Cartan subalgebra of the rank 16 10d gauge group, so that the Wilson lines yields $16(10-d)$ parameters. In total the number of background parameters is

$$(10-d)^2 + 16(10-d) = (26-d)(10-d), \quad (7.151)$$

in agreement with $\dim \mathcal{M}_d$.

Quantization in the Constant Background

We compactify the heterotic string on $(S^1)^{10-d}$ with periodic coordinates

$$x^m \simeq x^m + 2\pi R, \quad (7.152)$$

³⁵ But, of course, all lattices are equivalent over $\mathbb{R}$, so the phenomenon is pretty number theoretic.

and constant backgrounds for G_{mn} , B_{mn} , and A_m^I . Canonical quantization gives

$$k_{Lm} = \frac{n_m}{R} + \frac{w^n R}{\alpha'} (G_{mn} + B_{mn}) - q^I A_m^I - \frac{w^n R}{2} A_n^I A_m^I \quad (7.153)$$

$$k_L^I = (q^I + w^m R A_m^I) (2/\alpha')^{1/2} \quad (7.154)$$

$$k_{Rm} = \frac{n_m}{R} + \frac{w^n R}{\alpha'} (-G_{mn} + B_{mn}) - q^I A_m^I - \frac{w^n R}{2} A_n^I A_m^I \quad (7.155)$$

where n_m and w^m are *integers* (compact momenta and winding numbers) and q^I belongs to the Γ_{16} or to the $\Gamma_8 \oplus \Gamma_8$ lattice, depending of which 10d heterotic string we are compactifying.³⁶

Exercise 7.3 Derive Eqs. (7.153)–(7.155).

As a check, let us see that the momentum lattice Γ we get from canonical quantization, Eqs. (7.153)–(7.155), is indeed even and self-dual with respect to the natural indefinite pairing

$$(k_L^I, k_{Lm}, k_{Rm}) \circ (k_L'^I, k_{Lm}', k_{Rm}') \equiv k_L^I k_L'^I + G^{mn} (k_{Lm} k_{Ln}' - k_{Rm} k_{Rn}') \quad (7.156)$$

which is equal to (setting $\alpha' = 2$)

$$q^I q'^I + (n_m w'^m + n'_m w^m) \equiv (l_L^I, l_{Lm}, l_{Rm}) \circ (l_L'^I, l_{Lm}', l_{Rm}') \quad (7.157)$$

which is nothing else than the $\circ$ product in the even, self-dual lattice

$$Spin(32)^+ \oplus (10 - d) H \quad \text{or} \quad 2 E_8 \oplus (10 - d) H \quad (7.158)$$

where the notation is as in **Corollary 7.1**.

★ Geometry of Moduli Spaces at Infinity

The moduli spaces are non-compact, so we can find points at arbitrary large distance from a chosen reference point (corresponding to a lattice Γ_0). Points at infinite distance in the moduli space play a major role in the swampland program [36, 37], and their physics is of central importance in quantum gravity. In this section we give a watered-down description of the moduli geometry at infinite distance. To cover the toroidal compactifications of heterotic and bosonic strings, as well as of Type II strings (to be discussed in Sect. 13.2) and other important situations [8], we consider a general moduli space of the form

$$\mathcal{M} = G(\mathbb{Z}) \backslash \widetilde{\mathcal{M}}, \quad \text{with} \quad \widetilde{\mathcal{M}} = G(\mathbb{R})/K, \quad (7.159)$$

³⁶ **Beware:** for simplicity in this section we normalize the q^I so that the inner product takes the form $q^I q'^I$; in this normalization the charges are *not* integers.

where K is a maximal compact subgroup of the simple real Lie group³⁷ $G(\mathbb{R})$ and $G(\mathbb{Z}) \subset G(\mathbb{R})$ is a maximal arithmetic subgroup. For toroidal compactifications down to d dimensions³⁸

bosonic $G = O(26 - d, 26 - d)$
heterotic $G = O(26 - d, 10 - d)$
Type II G the $\mathbb{R}$ -split form of E_{11-d} (cf. Sect. 13.2)

The locally symmetric spaces (7.159) are equipped with their symmetric metric³⁹ (unique up to overall normalization). $\widetilde{M}$ is smooth, simply-connected, complete, non-compact, and has non-positive sectional curvatures. Then, by the Cartan-Hadamard theorem [26, 38], $\widetilde{M}$ is diffeomorphic to $\mathbb{R}^n$ via the exponential map $\exp_{x_0} : T_{x_0}\widetilde{M} \rightarrow \widetilde{M}$. Here x_0 is a chosen reference point.

The quotient M is not smooth since $G(\mathbb{Z})$ does not act freely (the orbifold singularities are points of gauge symmetry enhancement); however it has a *finite* cover $M' \rightarrow M$ which is smooth $M' = \Gamma \backslash \widetilde{M}$ where $\Gamma \triangleleft G(\mathbb{Z})$ is a finite-index normal subgroup and $\Gamma = \pi_1(M')$. M and M' have *finite volume* a fact with deep implications of physical relevance [39]. We are interested in the spaces $M(\infty)$ and $\widetilde{M}(\infty)$ of ‘points at infinite distance’ in these moduli spaces. For the full story we refer the reader to the deep books [40, 41] and references therein. Here we work in a cavalier fashion, glossing over subtleties and technicalities.

The notion of ‘point at infinity’ of an open space X may be made precise by introducing a compactification $\overline{X}$ of X : the space of points at infinity is then $X(\infty) = \overline{X} \setminus X$ with the induced topology. A natural way⁴⁰ to compactify a simply-connected, complete, non-compact Riemann manifold X with non-positive sectional curvatures is to consider the set of geodesics⁴¹ emanating from a reference point x_0 and declare two geodesics $\gamma_1(t)$, $\gamma_2(t)$ to be equivalent if $\lim_{t \rightarrow +\infty} d(\gamma_1(t), \gamma_2(t)) < \infty$. A point at infinity is then an equivalence class of geodesics. We write $X(\infty)$ for the set of points at infinity. The space $\overline{X} = X \cup X(\infty)$ (endowed with the appropriate topology) is the *geodesic compactification* of X [40]. Informally, each class of geodesics selects a direction along which we may reach infinity. The map $\exp_{x_0} : T_{x_0}X \rightarrow X$ identifies $X(\infty)$ with the unit sphere in the $T_{x_0}X$.

Exercise 7.4 Construct the geodesic compactification of $\mathbb{R}^n$ with the flat metric.

The geodesic compactification applies to the covering moduli $\widetilde{M} \equiv G(\mathbb{R})/K$. A crucial observation⁴² is that the isotropy group $I_p \subset G(\mathbb{R})$ of a point $p \in X(\infty)$ is a proper parabolic subgroup, and conversely each proper parabolic subgroup P is the isotropy group of some point in $X(\infty)$; in general the point is *non-unique*: P fixes point-wise some subset $X_P \subset X(\infty)$, and $X(\infty)$ is the disjoint union⁴³ of X_P over all proper parabolic subgroups of $G(\mathbb{R})$. Two points at infinity have the “same geometry” iff their parabolic groups are conjugate⁴⁴ in $G(\mathbb{R})$.

³⁷ In the text we are *very* sloppy. To be mathematically precise we have to define G as a simple (linear, connected) algebraic group defined over $\mathbb{Q}$, so that $G(\mathbb{R})$ is the real Lie groups of its $\mathbb{R}$ -valued points, while $G(\mathbb{Z}) \subset G(\mathbb{Q})$ is an arithmetic subgroup fixing a lattice $L \in \mathbb{Q}^n$ [24].

³⁸ It is of paramount importance that *physics* predicts a precise $\mathbb{Q}$ -algebraic structure for each one of these groups. $O(r, s) = \{g \in GL(\mathbb{Q}) : g^t \eta g = \eta\}$ where η is the diagonal matrix with $r + 1$ ’s and $s - 1$ ’s on the diagonal. E_{11-d} stands for the *universal Chevalley group of type* E_{11-d} .

³⁹ Recall that a metric is (locally) symmetric iff its Riemann tensor is covariantly constant $\nabla_i R_{jklm} = 0$. An irreducible symmetric Riemannian manifold is Einstein: $R_{ij} = \lambda G_{ij}$.

⁴⁰ But not the only one way.

⁴¹ All geodesics $\gamma(t) \subset X$ are parametrized by the arc-length t .

⁴² Cf. **Proposition I.2.6** of [40].

⁴³ Cf. **Proposition I.2.16** of [40]. $X(\infty)$ then acquires the structure of the *Tits building* of $G(\mathbb{R})$, cf. **Proposition I.2.19** of [40].

⁴⁴ Conjugacy classes of parabolic subgroups are in one-to-one correspondence with the subsets of the simple roots of Cartan’s *relative* root system of $G(\mathbb{R})/K$ (see e.g. Chap. 32 of [35]).

The story with the compactification of the arithmetic quotient $\mathcal{M} \equiv G(\mathbb{Z}) \backslash G(\mathbb{R}) / K$ is morally similar but technically more involved. The images of *almost all* geodesics of $\mathcal{M}$ under the covering map $\tilde{\mathcal{M}} \rightarrow \mathcal{M}$ are geodesics in $\mathcal{M}$ which *never* reach points at infinity in $G(\mathbb{Z}) \backslash G(\mathbb{R}) / K$: they wander around forever in some finite region. Only the few ones which manage to reach infinity⁴⁵ contribute to $\mathcal{M}(\infty)$. One should consider only the proper parabolic subgroups of the $\mathbb{Q}$ -algebraic group G which are defined over $\mathbb{Q}$: this gives the Tits $\mathbb{Q}$ -building of G , and then take the quotient by $G(\mathbb{Z})$ to get $\mathcal{M}(\infty)$. The number of $G(\mathbb{Z})$ -orbits of proper $\mathbb{Q}$ -parabolic groups is finite,⁴⁶ so we have only finitely many distinct ‘physical behaviours’ at infinity. A couple of examples is in order.

Example 1: $G = SO(2, 1) \simeq SL(2, \mathbb{R})$

The covering space $\mathcal{H} \equiv SO(2, 1)/SO(2)$ is the upper half-space $\mathcal{H}$ ($\equiv$ the unit disk $\mathcal{D}$); clearly $\mathcal{H}(\infty) = \mathbb{R} \cup i\infty$. The geodesics in $\mathcal{H}$ are either vertical lines $\tau(t)_{i\infty} = x + ie^t$, which form the equivalence class of the point $\tau = i\infty$ or semicircles with center on the real axis $\tau(t)_z = z(x + ie^t)/(x + ie^t + z)$ which form the equivalence class of $z \in \mathbb{R}$. The point $z \in \mathbb{R}$ is fixed by the parabolic group of elements of the form

$$P_z = \begin{pmatrix} a^{-1} - bz & (a - a^{-1})x + bz^2 \\ -b & a + bz \end{pmatrix} \subset SL(2, \mathbb{R}) \simeq SO(2, 1) \quad a \in \mathbb{R}_{>0}, b \in \mathbb{R}. \quad (7.160)$$

P_z is defined over $\mathbb{Q}$ iff $z \in \mathbb{Q} \cup i\infty$. Correspondingly, the image of the geodesics $\tau(t)_z$ in the moduli space $SL(2, \mathbb{Z}) \backslash \mathcal{H}$ is a EDM geodesic iff $z \in \mathbb{Q}$. Since $SL(2, \mathbb{Z})$ acts transitively on $\mathbb{Q} \cup i\infty$, the arithmetic quotient $SL(2, \mathbb{Z}) \backslash \mathcal{H}$ has a *single* point at infinity (i.e. only one cusp).

Example 2: decompactification limits in Narain moduli spaces

We see $O(k, k)$ as (here $\mathbf{1}_k$ is the $k \times k$ identity matrix)

$$O(k, k) = \{g \in GL(2k): g^t \Sigma g = \Sigma\}, \quad \Sigma = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \otimes \mathbf{1}_k \quad (7.161)$$

Let $L = \text{diag}(L_i)$ be a real diagonal $k \times k$ matrix. All geodesics in $O(k, k)/O(k)^2$ are equivalent to some $O(k, k)$ -translate of

$$\gamma(t) = \begin{pmatrix} \exp(tL) & 0 \\ 0 & \exp(-tL) \end{pmatrix} \quad (7.162)$$

for some L . Comparing with Eq. (★★) in **BOX 6.8** we see that the compactification radii evolve along the geodesic (7.162) as

$$R_i(t) = \exp(tL_i) \quad t \rightarrow +\infty \quad (7.163)$$

so that all points at infinity correspond to a certain number of radii going to infinity and/or zero. Taking the quotient by the action of the T -duality group, we may identify all points at infinity as points where some radii are sent to infinity i.e. as partial decompactification limits.

⁴⁵ Such geodesics are called EDM geodesics (eventually distance minimizing) cf Sect. III.20 of [40].

⁴⁶ Cf. **Proposition III.2.19** of [40]

BOX 7.4 - Proof of the Duality (7.169)

Clearly only the zero modes matter, so we forget the oscillators. The condition that the two states are in the same representation of $SO(16) \times SO(16)$ means $k'_L = k_L$ that is

$$q'^I + w' R' A'^I = q^I + w R A^I. \quad (\spadesuit)$$

where $q'^I \in \Gamma_8 \oplus \Gamma_8$ and $q^I \in \Gamma_{16}$. The condition $\sum_{I=1}^8 q'^I = 0 \bmod 2$ implies

$$w' = w' \sum_{I=1}^8 R' A'^I \equiv \sum_{I=1}^8 q^I + 4w \equiv \sum_{I=1}^8 q^I \bmod 2. \quad (\star)$$

On the other hand, the condition $q^1 = -q^9 \bmod 1$ implies

$$2(q'^1 + q'^9) = 2\left(q^1 + \frac{1}{2}w + q^9\right) - 2w' \equiv w \bmod 2. \quad (\star\star)$$

The momenta in the 9-direction are

$$k_{L,R} = \frac{1}{R} \left(n - w - \frac{1}{2} \sum_{i=1}^8 q^i \right) \pm \frac{wR}{\alpha'}, \quad k'_{L,R} = \frac{1}{R'} \left(n' - w' - (q'^1 + q'^9) \right) \pm \frac{w'R'}{\alpha'}$$

If R and R' are as in Eq. (7.167), the spectra are identified under $(k_L, k_R) \leftrightarrow (k'_L, -k'_R)$ iff

$$2(n - w) - \sum_{I=1}^8 q^I = w', \quad w = 2(n' - w') - 2(q'^1 + q'^9)$$

and $(\spadesuit)$ holds. Eqs. $(\star), (\star\star)$ imply that given $q^I \in \Gamma_{16}$ and $n, w \in \mathbb{Z}$ there exists a (unique) solution to these conditions with $q'^I \in \Gamma_8 \oplus \Gamma_8$ and $n', w' \in \mathbb{Z}$ (and viceversa).

If $P \subset G(\mathbb{Q})$ is a $\mathbb{Q}$ -parabolic subgroup, $|P \cap G(\mathbb{Z})| = \infty$, i.e. a point at infinity in the moduli space $G(\mathbb{Z}) \backslash G(\mathbb{R})/K$ is fixed by an *infinite* subgroup of the T -duality group. The bottom line is that the theory of points at infinity in moduli space is ‘morally’ analogue to the theory of special enhancement points, with infinite isotropy subgroup replacing the finite ones.

7.7.1 Relation Between $E_8 \times E_8$ and $SO(32)$ Heterotic Strings

For $d < 10$, the moduli space $\mathcal{M}_d$ is *connected*. This corresponds to the fact that any two even self-dual lattices in indefinite signature can be continuously deformed into each other. This implies that we get the same set of models whether we compactify the $SO(32)$ or the $E_8 \times E_8$ heterotic string. This yields a direct connection between

the two 10d heterotic strings which then should be considered as different phases (or vacua) of the same underlying theory.

As illustrated in **Example 2** above, a point at infinity in $\mathcal{M}_d$ corresponds physically to a limit where one takes some of the compact coordinates to infinite radius. There are several physically inequivalent limits in correspondence with the distinct $\mathbb{Q}$ -conjugacy classes of proper $\mathbb{Q}$ -parabolic subgroups. One particular limit produces the $SO(32)$ heterotic string in 10d and a different one the $E_8 \times E_8$ model. The $\mathbb{Q}$ -parabolic isotropy groups of these two limits are conjugate in $O(26 - d, 10 - d, \mathbb{R})$ but not in $O(26 - d, 10 - d, \mathbb{Q})$: from this point of view the existence of two 10d SUSY heterotic strings has a subtle Number-Theoretic origin.

We describe explicitly the continuous deformation connecting the two 10d models. Unprimed quantities refer to the $SO(32)$ description, and primed ones to the $E_8 \times E_8$ one. We compactify the $SO(32)$ model on a circle of radius R with $G_{99} = 1$ and Wilson line ($\equiv$ gauge holonomy)

$$RA_9^I = \text{diag}\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}, 0, 0, 0, 0, 0, 0, 0, 0\right). \quad (7.164)$$

The adjoint states with one index $1 \leq A \leq 16$ and one $17 \leq B \leq 32$ become antiperiodic due to the Wilson line, so the gauge symmetry is reduced to

$$SO(16) \times SO(16). \quad (7.165)$$

We compactify the $E_8 \times E_8$ theory on a S^1 of radius R' with $G'_{99} = 1$ and Wilson line

$$R'A'_9 = (1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0). \quad (7.166)$$

Now the states from the $SO(16)$ root lattice of each E_8 have integral weights and remain periodic, while the ones in the $SO(16)$ spinor lattice in each E_8 become anti-periodic. Again the unbroken gauge group is $SO(16) \times SO(16)$.

We claim that if we choose the two radii to be related as

$$RR' = \alpha'/2, \quad (7.167)$$

the two compactified theories become *physically equivalent*, i.e. dual descriptions of the *same* theory. To justify this assertion we need to show that they have the same spectrum in each sector of definite charge for the gauge group $SO(16) \times SO(16)$.

In the $SO(32)$ language the 9d mass formulae read ($k_L \equiv k_{L9}, k_R \equiv k_{R9}$)

$$\frac{1}{2}m^2 = \frac{1}{2} \sum_I k_L^I k_L^I + \frac{1}{2} k_L k_L + N - 1 = \frac{1}{2} k_R k_R + \tilde{N} - \nu \quad (7.168)$$

where the momenta (k_L^I, k_L, k_R) are given explicitly in Eqs. (7.153)–(7.155) in terms of the background, $q^I \in \Gamma_{16}$, winding number $w \in \mathbb{Z}$, and compact momentum $n \in \mathbb{Z}$. The same formulae apply to the primed background in the $E_8 \times E_8$ framework

with $q'^I \in \Gamma_8 \oplus \Gamma_8$. We claim that when (7.167) holds the two spectra are identical under the identification

$$(k_L'^I, k_L', k_R') = (k_L^I, k_L, -k_R) \quad (7.169)$$

i.e. under ‘parity on the right side only’. The first equality $k_L'^I = k_L^I$ expresses the fact that the states in the two sides of the duality transform the same way under $SO(16) \times SO(16)$, while the sign flip in the right momentum $k_R \leftrightarrow -k_R$ leaves their mass invariant. To get our claim it remains to show that the dual momenta $(k_L'^I, k_L', k_R')$ are still given by Eqs. (7.153)–(7.155) for the primed background with $q'^I \in \Gamma_8 \oplus \Gamma_8$ and $w', n' \in \mathbb{Z}$: see **BOX 7.4** for a detailed proof.

7.7.2 Example: Toroidal Compactification to Four Dimensions

As an example we consider compactification down to four dimensions. At generic points in the moduli space $\mathcal{M}_4$, the massless spectrum is given by the field-theoretic dimensional reduction of the 10d massless spectrum i.e. by simply decomposing the 10d massless spectrum into representations of the 4d Poincaré group.

Toroidal compactification does not reduce the number of supercharges: we start with 16 in 10d and we end up with 16 in 4d, that is $\mathcal{N} = 4$ SUSY.⁴⁷ Thus we get a 4d $\mathcal{N} = 4$ supergravity. The *generic* massless content is given by

- the 4d graviton $G_{\mu\nu}$;
- $16 + 6 + 6$ vectors. The first 16, A_μ^a , belong to the Cartan algebra of 10d gauge group left unbroken by the generic Wilson lines. In addition we have the KK vectors $G_{\mu m}$ and $B_{\mu m}$ with $m = 4, \dots, 9$;
- $(10 - 4)(26 - 4) = 132$ (real) scalars from moduli plus two ‘universal’ scalars: ϕ (from the 10d dilaton) and the 4d scalar field σ dual to the 4d 2-form $B_{\mu\nu}$ such that $d\sigma \equiv *_4 dB$;
- 4 Majorana gravitinos ψ_μ^A ($A = 1, 2, 3, 4$);
- $4(16 + 6 + 1)$ spin- $\frac{1}{2}$ fermions $\lambda^{aA}, \psi_m^A, \chi^A$.

The 4d $\mathcal{N} = 4$ gravity supermultiplet contains the graviton, 4 gravitinos, 6 vectors (called *graviphotons*), 4 spin- $\frac{1}{2}$, and one complex scalar τ . The only $\mathcal{N} = 4$ matter multiplet (the *vector multiplet*) contains a vector, 4 spinors, and 6 real scalars. Hence at a generic point we have the gravitational multiplet coupled to 22 vector multiplets.

At a point $p \in \mathcal{M}_4$ of enhanced gauge symmetry with gauge group

$$G = G_L \times U(1)^6, \quad (7.170)$$

⁴⁷ A different (and more fundamental) argument is given in footnote 18 of this chapter.

we have $\dim G_L$ vector multiplets transforming in the adjoint of G_L . The $U(1)^6$ vectors with vertices $\partial X_\mu \tilde{\psi}^m e^{-\tilde{\phi}} e^{ikX}$ are then the graviphotons.

The scalar sector of 4d $\mathcal{N} = 4$ SUGRA coupled to n vector multiplets is described by a σ -model with target-space the quotient of the symmetric space [8, 42, 43]

$$\frac{SU(1, 1)}{U(1)} \times \frac{SO(n, 6)}{SO(n) \times SO(6)} \quad (7.171)$$

by a discrete subgroup Γ of the isometry group $SU(1, 1) \times SO(6, n)$. Considerations of consistency of quantum gravity lead to the requirement that Γ is an *arithmetic* subgroup, see [8] for a complete discussion. The first factor in (7.171) is the upper half-plane $\mathcal{H}$; it is the target space for the complex scalar τ in the gravitational supermultiplet which combines the two universal scalars: one has $\tau = \sigma + ie^{-\phi}$. The scalars which are moduli of the toroidal compactification then make the second factor space in (7.171). Taking $n = 22$, this second factor space becomes precisely the covering moduli space $\widetilde{\mathcal{M}}_4$ in Eq. (7.134). We see that the stringy result (7.134) could be deduced using four-dimensional supersymmetry only.

The new information, which is purely stringy, is that we have to take the quotient by $SO(n, 6; \mathbb{Z})$, i.e. that the T -duality group acts as a group of *gauge* symmetries. We expect that Γ is an arithmetic subgroup of the full isometry group, hence commensurable to $SL(2, \mathbb{Z}) \times SO(n, 6; \mathbb{Z})$. The factor $SL(2, \mathbb{Z})$ should correspond to a new duality called *S-duality*. However this putative duality would map the string coupling constant $g = e^\Phi \rightarrow e^{-\Phi} = g^{-1}$ so it would relate weak coupling to strong coupling and cannot be visible in string perturbation theory. To address the issue of the existence of this putative $SL(2, \mathbb{Z})$ *S-duality* we need to wait the discussion of strings beyond perturbation theory in Chap. 13.

★ Low-Energy Scalars' Target-Spaces

In addition to the “matter” vectors (7.185), we have the $10 - d$ vectors with vertices (7.138) where the Lorentz vector index μ is carried by a left-mover. From the point of view of SUSY in d dimensions, these are *graviphotons* that is, massless vectors which belong to the gravity supermultiplet (i.e. the supermultiplet containing the metric field $g_{\mu\nu}$ and the gravitinos). Just as in the $d = 4$ case, Eq. (7.171), the universal cover of the 16-supercharges d -dimensional SUGRA scalars' target-space, $\widetilde{\mathcal{M}}_d$, factors in the cover of the lattice moduli space $\widetilde{\mathcal{M}}_d$ in (7.135) and the target-space for the *graviscalars* (scalars in the gravity supermultiplet). In $d \geq 5$ the only graviscalar is the dilaton and

$$d \geq 5 \quad \widetilde{\mathcal{M}}_d = \mathbb{R} \times \widetilde{\mathcal{M}}_d, \quad (7.172)$$

while in $d = 4$ we have Eq. (7.171). $d = 3$ is special since the metric $g_{\mu\nu}$ does not propagate any local degree of freedom (no gravitational-waves in 3d), so all local d.o.f. belong to ‘matter’ supermultiplets, and hence the lattice moduli and all the other scalars should unify in a single irreducible covering space. Moreover in 3d massless vectors are dual to scalars, and $\widetilde{\mathcal{M}}_d$ contains these dual scalars as well; the scalars' space of the 3d 16-supercharge supergravity then has dimension

$$\overbrace{(26-3)(10-3)}^{\text{lattice moduli}} + \overbrace{(26-3) + (10-3)}^{\text{dual generic vectors}} + \overbrace{1}^{\text{dilaton}} = 8 \cdot 24 \quad (7.173)$$

and one has [8, 44, 45]

$$d = 3 \quad \widetilde{\mathcal{M}}_3 = O(24, 8) / [O(24) \times O(8)] \quad (7.174)$$

isometric to $\widetilde{\mathcal{M}}_2$.

7.8 Supersymmetry and BPS States

As we shall show momentarily, the SUSY algebra of the toroidally compactified heterotic theory has the form

$$\{Q_\alpha, Q^\beta\} = 2P_\mu (\Gamma^\mu \Gamma^0)_\alpha{}^\beta + 2P_{Rm} (\Gamma^m \Gamma^0)_\alpha{}^\beta \quad (7.175)$$

where α, β are indices of the $\mathbf{16}_s$ of $\text{Spin}(9, 1)$. Equation (7.175) differs from the obvious dimensional reduction of the 10d SUSY algebra by the replacement of the compact momentum operator P_m by its right part P_{Rm} , whose eigenvalue is the *total right-moving momentum* k_{Rm} of all strings in the given state. P_m and P_{Rm} are equal only for states of total winding number zero. For definiteness, we consider heterotic strings compactified to $d = 4$; from the viewpoint of the 4d SUSY algebra, P_{Rm} are SUSY *central charges* [46].

To see that (7.175) is the correct 4d SUSY algebra is very easy: we use the world-sheet $(0, 1)$ currents which generate space-time supersymmetry (cf. Sect. 3.8). In the heterotic string we have only right-moving $\text{Spin}(9, 1)$ spin-fields, and so the $(0, 1)$ currents which generate space-time SUSY are

$$\text{picture } -\frac{1}{2}: \quad Q_\alpha^{(-1/2)}(\bar{z}) \equiv \tilde{S}_\alpha(\bar{z}) e^{-\tilde{\phi}(\bar{z})/2} \quad (7.176)$$

$$\text{picture } +\frac{1}{2}: \quad Q_\alpha^{(+1/2)}(\bar{z}) \equiv \left(\bar{\partial} X_\mu \Gamma_{\alpha\dot{\beta}}^\mu + \bar{\partial} X_m \Gamma_{\alpha\dot{\beta}}^m \right) \tilde{S}^{\dot{\beta}}(\bar{z}) e^{\tilde{\phi}(\bar{z})/2}, \quad (7.177)$$

with OPE

$$Q_\alpha^{(-1/2)}(\bar{z}) Q_\beta^{(+1/2)}(\bar{w}) \sim \frac{1}{z-w} \left(\bar{\partial} X_\mu (\Gamma^\mu \Gamma^0)_{\alpha\beta} + \bar{\partial} X_m (\Gamma^m \Gamma^0)_{\alpha\beta} \right). \quad (7.178)$$

The charge associated to the current $\bar{\partial} X_\mu$ is the non-compact momentum $P_\mu \equiv P_{L\mu} \equiv P_{R\mu}$, while the charge associated to $\bar{\partial} X_m$ is P_{Rm} , as it was to be shown.

BPS States

We look for states which are annihilated by some of the supercharges Q_α : they are called *BPS states*.⁴⁸ Take the expectation value of both sides of (7.175) in any

⁴⁸ After Bogomol'nyi, Prasad, and Sommerfield [47–49].

single-string state $|\psi\rangle$ of mass M in its rest frame. The LHS is a non-negative matrix. The RHS is

$$2(M + k_{Rm} \Gamma^m \Gamma^0)_\alpha^\beta. \quad (7.179)$$

Since

$$(k_{Rm} \Gamma^m \Gamma^0)^2 = k_R^2, \quad (7.180)$$

the eigenvalues of the matrix (7.179) are

$$2(M \pm |k_R|) \geq 0 \quad \Rightarrow \quad M \geq |k_R| \quad (7.181)$$

with equal number of plus and minus signs. The inequality (7.181) is known as the *BPS bound*. A state that saturates it, i.e. of mass $M^2 \equiv k_R^2$, is a BPS state. Indeed saturating states are zero eigenvectors of the matrix (7.179), hence are annihilated by half the supercharges (8 out of 16).

Recalling the mass-shell condition on the right-hand side

$$M^2 = \begin{cases} k_R^2 + 4(\tilde{N} - 1/2)/\alpha' & \text{(NS)} \\ k_R^2 + 4\tilde{N}/\alpha' & \text{(R)}, \end{cases} \quad (7.182)$$

we see that the BPS states are the ones for which the right-movers are either a R ground state or a NS state with just one $\tilde{\psi}_{-1/2}$ excited, i.e. the lowest NS states which survive the GSO projection: by abuse of language we shall call ‘ground states’ both kinds of states. With this linguistic convention, the BPS states are precisely those states for which the right-moving side is in its $\mathbf{8}_v \oplus \mathbf{8}_s$ ‘ground state’ *but have an arbitrary large k_R* . These right-moving states can be paired with many possible states on the left-moving side. The left-mover mass-shell condition is

$$M^2 = k_L^2 + \frac{4}{\alpha'}(N - 1), \quad (7.183)$$

i.e.

$$N = 1 + \alpha'(k_R^2 - k_L^2)/4 = 1 - n \cdot w - \frac{1}{2}q^I q^I, \quad (7.184)$$

where in the last equality we used (7.157). Any left-moving oscillator state is consistent with the BPS condition, as long as the momenta, winding and charges q^I satisfy the relation (7.184). For any given such left-moving state, the 16 right-moving states in the $\mathbf{8}_v \oplus \mathbf{8}_s$ form an *ultrashort* representation of the SUSY algebra with $16 = 2^{16/4}$ states which is much shorter than the standard massive supermultiplet which has $256 = 2^{16/2}$ states.

Note that the “matter” gauge vectors and gluinos also form a ultrashort representation (they are 16 states per generator of the gauge group): the matter vector multiplets are the special BPS states with

$$k_R = 0, \quad \begin{cases} k_L^2 = \frac{2}{\alpha'} l_L^2 \equiv \frac{4}{\alpha'}, & N = 0 \\ k_L = 0, & N = 1. \end{cases} \quad (7.185)$$

10d Origin of the Modified SUSY Algebra

We may rewrite the algebra (7.175) in the form ($M = 0, \dots, 9$)

$$\{Q_\alpha, Q_\beta^\dagger\} = 2P_M(\Gamma^M \Gamma^0)_{\alpha\beta} - 2\frac{\Delta X_m}{2\pi\alpha'}(\Gamma^m \Gamma^0)_{\alpha\beta}, \quad (7.186)$$

where ΔX_m is the total winding of the strings. The central charge term in the supersymmetry algebra must be proportional to a conserved charge (by the HLS theorem [46]), so we are looking for a charge proportional to the length ΔX of a string. We already know such a charge: we showed in BOX 6.6 that the conserved charge which couples to the 2-form field B is precisely the total winding number

$$Q^M = \frac{1}{2\pi\alpha'} \int_{\ell_i} dX^M, \quad (7.187)$$

which is exactly the central charge entering in the SUSY algebra (7.186):

$$\{Q_\alpha, Q_\beta^\dagger\} = 2(P_M - Q_M)(\Gamma^M \Gamma^0)_{\alpha\beta}. \quad (7.188)$$

In 10 non-compact dimensions the charge (7.187) vanishes for any finite closed string, but it can be carried by an infinite string, for example an infinite straight string which would arise as the $R \rightarrow \infty$ limit of a winding string. Under compactification the combination $(P_m - Q_m)$ becomes the right-moving gauge charge P_{Rm} . We stress that the left-moving charges never appear in the SUSY algebra as part of the SUSY central charges. This reflects the fact that the corresponding gauge vectors are ‘matter’ fields, *not* graviphotons. Indeed in supergravity all generators of the SUSY algebra, momenta, supercharges, and central charges are expressed (in an asymptotically flat background) as fluxes at infinity of suitable ‘gauge field strengths’. Roughly speaking, the central charges are the flux at infinity of the graviphoton field strengths (see Chap. 6 and Sect. 4.7 of [8]). The electro-magnetic charges which couple to the ‘matter’ vectors cannot appear in the SUSY central charges.

Appendix: The $\mathcal{N} = 2$ Superstring

We briefly comment on the superstrings based on $\mathcal{N} = 2$ superconformal algebras. For simplicity we assume left-right symmetry, so we have a (2,2) superconformal algebra of gauge constraints. We focus on the very basic construction. For more general models see the original papers [50–52].

We consider first the case where the ‘matter’ is free as in Eq. (2.397). (2, 2) SUSY in 2d may be seen as the dimensional reduction of $\mathcal{N} = 1$ in 4d. In 4d $\mathcal{N} = 1$ SUSY [53] the basic matter superfield is the chiral one whose bosonic d.o.f. form a *complex* scalar Z and the fermionic ones a

4d Majorana spinor which becomes a Dirac spinor in 2d. Each free 2d chiral superfield contributes $2 \cdot 1 + 2 \cdot \frac{1}{2} = 3$ to c_{matter} . The ghosts' central charge for $N = 2$ is $c_{\text{ghost}} = -6$ (cf. Table 2.1), so we construct a *critical* $(2, 2)$ superstring by taking as matter two free chiral superfields (here $a = 1, 2$)

$$S = \frac{1}{2\pi} \int d^2z \left(\partial \bar{Z}^a \bar{\partial} Z_a + \bar{\psi}^a \bar{\partial} \psi_a + \bar{\tilde{\psi}}^a \partial \tilde{\psi}_a \right) \quad (7.189)$$

with left-moving superconformal currents

$$T = -\partial \bar{Z}^a \partial Z_a - \frac{1}{2} (\bar{\psi}^a \partial \psi_a + \psi_a \partial \bar{\psi}^a) \quad J = -\bar{\psi}^a \psi_a \quad (7.190)$$

$$G^+ = \sqrt{2}i \psi_a \partial \bar{Z}^a \quad G^- = \sqrt{2}i \bar{\psi}^a \partial Z_a \quad (7.191)$$

and the corresponding right-moving ones. The real dimension of the target space is 4, and the model has 4d translational invariance. However it is *not* 4d Poincaré invariant, since the fields come in complex pairs and only rotations commuting with their complex structure—i.e. rotations preserving the $U(1)$ current $J(z)$ —are symmetries of the gauge constraints; hence the rotational symmetry is $U(2)$ in Euclidean signature that may be ‘Wick rotated’ to $U(1, 1)$ which is a subgroup of $SO(2, 2)$. The target metric is either Euclidean or has signature $(2, 2)$ but is never Lorentzian.

Historically two different interpretations were offered for this string model. The older one saw the theory as being a Poincaré invariant theory in two dimensions; only the real parts of the Z_a were interpreted as space-time coordinates, while the imaginary parts were seen as internal degrees of freedom. A more natural and deeper interpretation [50–52] sees the model as describing a “physical” spacetime $\mathbb{C}^{1,1}$ of complex signature $(1, 1)$ i.e. $(2, 2)$ in real terms. Of course, a theory with two physical times sounds a bit exotic, but it is fully consistent at the mathematical level.

Let us describe the situation in a perhaps simpler language. We already said that 2d $(2, 2)$ SUSY is the dimensional reduction of 4d $N = 1$. We know that in 4d SUSY the target space must be a Kähler manifold $\mathcal{K}$ [54]; this result remains valid under dimensional reduction to 2d [55]. Then the $(2, 2)$ superstring should describe the motion of the string in a (pseudo)-Kähler space of complex dimension 2. In real terms a pseudo-Kähler metric has signature $(4, 0)$ or $(2, 2)$ but never the Lorentzian signature $(3, 1)$.⁴⁹ A (pseudo)-Kähler metric is a very special kind of Riemannian metric: it is Hermitean, that is, $g_{ab} = g_{\bar{a}\bar{b}} = 0$, and satisfies differential conditions

$$\partial_a g_{b\bar{c}} = \partial_b g_{a\bar{c}}, \quad \partial_{\bar{a}} g_{b\bar{c}} = \partial_{\bar{c}} g_{b\bar{a}}. \quad (7.192)$$

As long as our 2d theory is required to have $N \leq 1$ supersymmetry, we are free to choose *any* Riemann metric $g_{\mu\nu}$ as the σ -model target metric [55]. Then in $N \leq 1$ string theory the world-sheet coupling $g_{\mu\nu}$ is identified with the *off-shell* dynamical metric i.e. with the spacetime gravitational d.o.f. Of course, to correspond to an *on-shell* background, the metric $g_{\mu\nu}$ should satisfy the target Einstein equations $R_{\mu\nu} - \dots = 0$ which, from the world-sheet perspective, is the requirement that the metric beta-function vanishes, $\beta_{\mu\nu} = 0$, as required by 2d conformal invariance or rather by nilpotency of the BRST charge, $Q^2 = 0$, cf. Sect. 1.8. In the $N = 2$ case, *even off-shell* we are not free to choose any target metric: the off-shell metric should be Kähler. The spacetime gravitational degrees of freedom are not longer there: the only remaining freedom (locally in target-space) is to choose a Kähler potential K such that

$$g_{a\bar{b}} = \partial_a \partial_{\bar{b}} K. \quad (7.193)$$

The target-space ‘scalar’ field K is the only ‘gravitational’ degree of freedom which is in the game. Clearly, K cannot propagate 4d gravitons, since it has spin 0, while gravitons have spin 2. Yet K

⁴⁹ From [8] we know that in Lorentzian signature the only possible *reductive* holonomy group is $SO(d-1, 1)$, so no special holonomy may appear for Lorentzian spacetime unless they are generalized pp -waves which have holonomy Lie algebras with non-trivial radicals.

is a promising massless d.o.f. propagating in the 2-dimensional complex target space. It is hard to see anything else propagating in such a target space.

In the simplest version of the $\mathcal{N} = 2$ string the only physical state which *propagates* in spacetime is this massless scalar. Other modular invariant ‘GSO projections’ exists; they produce additional models. All these models describe a *finite* number of particles propagating in space-time, all massless [50–52]. The basic reason for this fact is easily understood from the light-cone quantization. Since time now is a complex line, the time-light coordinates $Z^\pm$ are complex and span a *complex* dimension 2 space $\mathbb{C}^{1,1}$; given that the *full* target space is $\mathbb{C}^{1,1}$, the *transverse space* is *zero-dimensional*, and hence the light-cone gauge kills *all* oscillator modes, leaving only the zero-modes of the longitudinal coordinates $Z^\pm$ which can describe only massless scalars.

In the simplest theory the only surviving state is the “vacuum” i.e. the appropriate Bose/Sea for the ghosts, which is easy to infer from the $\mathcal{N} = 1$ case, just remembering that now we have twice as many commuting ghosts, β_+, γ_+ and β_-, γ_- . The obvious physical vertex in the standard picture is

$$V(k_a, \bar{k}^a, z, \bar{z}) = c(z) \tilde{c}(\bar{z}) e^{-\phi_+(z) - \phi_-(z) - \tilde{\phi}_+(\bar{z}) - \tilde{\phi}_-(\bar{z})} e^{i\bar{k}^a Z_a(z, \bar{z}) + i k_a \bar{Z}^a(z, \bar{z})} \quad (7.194)$$

whose Virasoro weight is

$$h = -1 + 2 \cdot \frac{1}{2} + \bar{k}^a k_a \equiv \bar{k}^a k_a, \quad (7.195)$$

so that the on-shell condition is $\bar{k}^a k_a = 0$ and the particle is massless as expected.⁵⁰ Let us show that the vertex $V(k_a, \bar{k}^a, z, \bar{z})$ describes the quanta of the target field K (the Kähler potential). Recall from Chap. 3 that the picture-zero vertex is the integral in superspace of the matter factor of the picture -1 vertex; hence the integrated version of vertex (7.194) in picture zero corresponds to the superspace integral

$$V(k_a, \bar{k}^a) = \int d^2 z d^2 \theta d^2 \bar{\theta} e^{i\bar{k}^a Z_a + i k_a \bar{Z}^a}, \quad (7.196)$$

where Z_a (resp. $\bar{Z}^a$) is the full 2d chiral superfield (resp. antichiral). Let us consider the generating functional of the multiple correlation functions of the operators $V(k_a, \bar{k}^a)$

$$Z[\hat{\Phi}] \equiv \left\langle \exp \left[- \int d^4 k \hat{\Phi}(k_a, \bar{k}^a) V(k_a, \bar{k}^a) \right] \right\rangle = \left\langle \exp \left[- \int d^2 z d^2 \theta d^2 \bar{\theta} \Phi(Z_a, \bar{Z}^a) \right] \right\rangle, \quad (7.197)$$

where $\Phi(z_a, \bar{z}^a)$ is the Fourier transform of $\hat{\Phi}(k_a, \bar{k}^a)$

$$\Phi(z_a, \bar{z}^a) = \int d^4 k \hat{\Phi}(k_a, \bar{k}^a) e^{i k_b \bar{z}^b + i \bar{k}^b z_b}. \quad (7.198)$$

From Eq. (7.197) it is obvious that $Z[\hat{\Phi}]$ is just the partition function of the 2d supersymmetric σ -model with a shifted Kähler potential

$$K_0(Z_a, \bar{Z}^a) \longrightarrow K(Z_a, \bar{Z}^a)' = K_0(Z_a, \bar{Z}^a) + \Phi(Z_a, \bar{Z}^a), \quad (7.199)$$

which proves our claim that the spacetime field of the stringy massless particle (7.194) is the Kähler potential.

Let us discuss the dynamical equations of motion which describe the propagation in target space of the field K . Again they are dictated by the (super)conformal invariance of the world-sheet theory. At the one-loop level, the metric β -function of the σ -model is just the Ricci tensor, so, to this order, the e.o.m. just require the Kähler metric $\partial_a \partial_{\bar{b}} K$ to be *Ricci flat*, that is, a *Calabi-Yau metric*⁵¹

⁵⁰ Of course, the kinematics of massless particles in signature $(2, 2)$ is rather different from the usual Lorentzian one.

⁵¹ For the moment, Calabi-Yau metric is just synonym of a metric which is both *Kähler* and *Ricci flat*. In Chap. 11 we shall discuss this important geometry in detail.

$$R_{a\bar{b}} = 0. \quad (7.200)$$

One can show⁵² that in complex dimension 2 a metric is Calabi-Yau if and only if it is *hyper-Kähler*. Indeed, both conditions are equivalent to the statement that Riemann tensor 2-form $R^{a\bar{b}} \equiv R^{a\bar{b}}{}_{c\bar{d}} dz^c \wedge d\bar{z}^{\bar{d}}$ is anti-selfdual

$$R^{a\bar{b}} = - * R^{a\bar{b}}. \quad (7.201)$$

On the other hand, a SUSY σ -model with a hyperKähler metric has an enhanced global $\mathcal{N} = 4$ supersymmetry [8]. Higher loop corrections cannot spoil global supersymmetry, and so the σ -model metric should remain hyperKähler, hence Ricci flat, to all orders, that is, Eq. (7.200) are the *exact* target space equations of motion: **the $\mathcal{N} = 2$ string describes self-dual gravity in 4d.**

Thus the (simplest version) of the $\mathcal{N} = 2$ string is precisely equivalent to the *quantum* self-dual four-dimensional gravity, with no other degree of freedom. Since the spacetime spectrum of this model consists of just one local scalar field K , *naively* one would think natural to see this system as a mere *field theory* rather than a *full-fledged* string theory. But it is not so: self-duality gravity is rather hard to quantize as a *field theory*; the $\mathcal{N} = 2$ string provides a fully consistent quantization.

In terms of the field Φ , the e.o.m. take the form of the *complex Monge-Ampère equations* [56]

$$\det(\partial_a \partial_{\bar{b}} K_0 + \partial_a \partial_{\bar{b}} \Phi) = 1, \quad (7.202)$$

or in the differential form notation

$$(i\partial\bar{\partial}K)^2 = \varepsilon \wedge \bar{\varepsilon}, \quad (7.203)$$

where ε is a holomorphic $(2, 0)$ form.

References

1. Gross, D.J., Harvey, J.A., Martinec, E.J., Rohm, R.: Heterotic string theory. 1. The free heterotic string. Nucl. Phys. **B256**, 253 (1985)
2. Gross, D.J., Harvey, J.A., Martinec, E.J., Rohm, R.: Heterotic string theory. 2. The interacting heterotic string. Nucl. Phys. **B267**, 75–124 (1986)
3. Neitzke, A., Vafa, C.: Topological strings and their physical applications. [arXiv:hep-th/0410178](https://arxiv.org/abs/hep-th/0410178)
4. Bershadsky, M., Cecotti, S., Ooguri, H., Vafa, C.: Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes. Commun. Math. Phys. **165**, 311–428 (1994). [arXiv:hep-th/9309140](https://arxiv.org/abs/hep-th/9309140)
5. Green, M.B., Schwarz, J.H.: Covariant description of superstrings. Phys. Lett. B **136**, 367 (1984)
6. Adams, J.F.: Lectures on Exceptional Lie Groups. University of Chicago Press, Chicago (1996)
7. Fuchs, J., Schweigert, C.: Symmetries. Cambridge Monographs on Mathematical Physics, Lie Algebras and Representations. A Graduate course for Physicists. Cambridge University Press, Cambridge (1997)
8. Cecotti, S.: Supersymmetric Field Theories. Geometric Structures and Dualities. Cambridge University Press, Cambridge (2015)
9. Gorbatsevich, V.V., Onishchik, A.L., Vinberg, E.B.: Structure of Lie groups and Lie algebras. In: Lie Groups and Lie Algebras III. Encyclopaedia of Mathematical Sciences, vol. 41. Springer, Berlin (1994)
10. Conway, J.H., Sloane, N.J.A.: Sphere Packings, Lattices and Groups, 3rd edn. Springer, Berlin (1999)

⁵² See Chap. 11 for details.

11. Bourbaki, N.: *Éléments de mathématique. Groupes et algèbres de Lie. Chapitre 4 à 6.* Springer, Berlin (2007)
12. Alvarez-Gaume, L., Ginsparg, P.H., Moore, G.W., Vafa, C.: An $O(16) \times O(16)$ heterotic string. *Phys. Lett. B* **171**, 155–162 (1986)
13. Lang, S.: *Algebra.* Springer, Berlin (2002)
14. Narain, K.S.: New heterotic string theories in uncompactified dimensions < 10 . *Phys. Lett. B* **169**, 41–46 (1986)
15. Narain, K.S., Sarmadi, M.H., Witten, E.: A note on toroidal compactification of heterotic string theory. *Nucl. Phys. B* **279**, 369–379 (1987)
16. Serre, J.-P.: *A Course in Arithmetics.* Springer, Berlin (1973)
17. NIST Digital Library of Mathematical Functions. <https://dlmf.nist.gov>
18. Serre, J.-P.: *Complex Semisimple Lie Algebras.* Springer, Berlin (2001)
19. Helgason, S.: *Differential Geometry. Lie Groups and Symmetric Spaces.* Academic, Cambridge (1978)
20. Milne, J.S.: *Algebraic Groups, The Theory of Group Schemes of Finite Type over a Field.* Cambridge University Press, Cambridge (2017)
21. Milne, J.S.: *Algebraic Groups, Lie Groups, and their Arithmetic Subgroups*, available from the author webpage <https://www.jmilne.org>
22. Milne, J.S.: *Lie Algebras, Algebraic Groups, and Lie Groups*, available from the author webpage <https://www.jmilne.org>
23. Borel, A.: *Introduction aux groupes arithmétiques.* Hermann, Missouri (1969)
24. Morris, D.W.: *Introduction to Arithmetic Groups.* [arXiv:math/0106063](https://arxiv.org/abs/math/0106063)
25. Ji, L.: *Arithmetic Groups and Their Generalizations. What, Why, and How.* AMS/IP Studies in Advanced Mathematics, vol. 43. AMS, Providence (2008)
26. Postnikov, M.M.: *Geometry VI. Riemannian Geometry.* Encyclopaedia Mathematica Sciences, vol. 91. Springer, Berlin (2001)
27. Serre, J.-P.: Bounds for the orders of the finite subgroups of $G(k)$. In: Geck, M., Testerman, D., Thévenaz, J. (eds.) *Group Representation Theory.* EPFL Press, Lausanne (2006)
28. Shephard, G.C., Todd, J.A.: Finite unitary reflection groups. *Can. J. Math.* **6**, 274–304 (1954)
29. Dolgachev, I.G.: *Reflection Groups in Algebraic Geometry.* Bulletin of the AMS, Providence (2008)
30. Geck, M., Malle, G.: *Contribution to the Handbook of Algebra: Reflection Groups.* [arXiv:math/0311012](https://arxiv.org/abs/math/0311012)
31. Lehrer, G.I., Taylor, D.E.: *Unitary Reflection Groups.* Cambridge University Press, Cambridge (2009)
32. Broué, M., Malle, G., Roquier, R.: Complex reflection groups, braid groups. Hecke algebras. *J. Reine Angew. Math.* **500**, 127–190 (1998)
33. Broué, M.: *Introduction to Complex Reflection Groups and Their Braid Groups.* Lecture Notes in Mathematics, vol. 1988. Springer, Berlin (2010)
34. Broué, M.: *Reflection Groups, Braid Groups, Hecke Algebra, Finite Reductive Groups, Current Developments in Mathematics (2000), pp. 1–107.* International Press, Vienna (2001)
35. Bump, D.: *Lie Groups. Graduate Text in Mathematics, vol. 225.* Springer, Berlin (2004)
36. Ooguri, H., Vafa, C.: On the geometry of the string landscape and the swampland. *Nucl. Phys. B* **766**, 21–33 (2007). [arXiv:hep-th/0605264](https://arxiv.org/abs/hep-th/0605264)
37. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](https://arxiv.org/abs/1711.00864)
38. Jöst, J.: *Riemannian Geometry and Geometric Analysis, 7th edn.* Springer, Berlin (2007)
39. Cecotti, S.: Swampland geometry and the gauge couplings. *JHEP* **09**, 136 (2021). [arXiv:2102.03205](https://arxiv.org/abs/2102.03205)
40. Borel, A., Ji, L.: *Compactifications of Symmetric and Locally Symmetric Spaces.* Birkhäuser (2005)
41. Guivarc’h, Y., Ji, L., Taylor, J.C.: *Compactifications of Symmetric Spaces.* Birkhäuser (1998)
42. Bergshoeff, E., Koh, I.G., Sezgin, E.: Coupling Yang-Mills to $N = 4$ $d = 4$ supergravity. *Phys. Lett. B* **155**, 71–75 (1985)

43. de Roo, M.: Matter coupling in $N = 4$ supergravity. Nucl. Phys. B **255**, 515–531 (1985)
44. de Wit, B., Tollsten, A.K., Nicolai, H.: Locally supersymmetric $D = 3$ nonlinear sigma models. Nucl. Phys. B **392**, 3–38 (1993)
45. de Wit, B., Berger, I., Samtleben, H.: Gauged locally supersymmetric $D = 3$ nonlinear sigma models. Nucl. Phys. B **671**, 175–216 (2003)
46. Haag, R., Łopuszanski, J.T., Sohnius, M.: All possible generators of supersymmetries of the S -matrix. Nucl. Phys. B **88**, 257 (1975)
47. Bogomolnyi, E.B.: The stability of classical solutions, Sov. J. Nucl. Phys. (Engl. Transl.) **24**, 449 (1976)
48. Prasad, M.K., Sommerfield, C.M.: Exact classical solution for the 't Hooft monopole and the Julia-Zee dyon. Phys. Rev. Lett. **35**, 760–762 (1975)
49. Manton, N., Sutcliffe, P.: Topological Solitons. Cambridge University Press, Cambridge (2004)
50. Ooguri, H., Vafa, C.: Selfduality and $N = 2$ string magic. Mod. Phys. Lett. A **5**, 1389–1398 (1990)
51. Ooguri, H., Vafa, C.: Geometry of $N = 2$ strings. Nucl. Phys. B **361**, 469–518 (1991)
52. Ooguri, H., Vafa, C.: $N = 2$ heterotic strings. Nucl. Phys. B **367**, 83–104 (1991)
53. Wess, J., Bagger, J.: Princeton Series in Physics. Supersymmetry and Supergravity. Princeton University Press, Princeton (1983)
54. Zumino, B.: Supersymmetry and Kahler manifolds. Phys. Lett. B **87**, 203 (1979)
55. Alvarez-Gaume, L., Freedman, D.Z.: Geometrical structure and ultraviolet finiteness in the supersymmetric sigma model. Commun. Math. Phys. **80**, 443 (1981)
56. Yau, S.T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampere equation. I. Commun. Pure Appl. Math. **31**, 339–411 (1978)

Part III

Physics of Supersymmetric Strings

In Part III, we study the physics of the supersymmetric string theory constructed in Part II at weak coupling or low-energy. Part III consists of four chapters.

In Chap. 8, we study the low-energy limit of the various supersymmetric string theories using techniques and results from supersymmetry and supergravity.

In Chap. 9, we show that the SUSY string theories are anomaly-free first by direct computation of the anomalies, and then by showing that modular invariance implies absence of anomalies.

In Chap. 10, we compute some sample amplitudes for the superstring to illustrate the techniques, deduce some general results as the non-renormalization theorems, and check some of the properties predicted in Chap. 9.

In Chap. 11, we introduce and study Calabi–Yau compactifications from a number of different viewpoints.

Chapter 8

Low-Energy Effective Theories



Abstract We construct the low-energy effective theories of the five 10d supersymmetric strings: their Lagrangians are essentially determined by SUSY. The effective theories yield a description which is non-perturbative in the sense that, while it is valid only asymptotically for small momenta, in this IR regime remains reliable for *all* values of the string coupling constant g . This remarkable property reflects the existence of powerful SUSY non-renormalization theorems for the relevant couplings and their dependence on g . We construct the half-BPS solutions of the effective theories: here the fundamental fact is that the SUSY central charges are not renormalized, so the results we get at the effective level are *exact* in the full string theory.

The 10d supersymmetric string theories either have $n_s = 32$ supercharges (Types IIA & IIB) or $n_s = 16$ supercharges (Type I & heterotic). Their low-energy limit is then a supersymmetric extension of Einstein gravity i.e. a *supergravity* (SUGRA). It is well known [1] that a SUGRA with more than 8 supercharges is determined by symmetry considerations up to a few global issues. This observation applies in particular to the low-energy limit of the SUSY strings.

Supergravity was introduced in [2, 3]; general references are [1, 4–7].

8.1 Supergravity: a Quick Review

By a *supergravity* (SUGRA) we mean a SUSY theory in $d \geq 3$ dimensions which contains Einstein gravity, i.e. includes a metric field $g_{\mu\nu}$ whose e.o.m. have the form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = T_{\mu\nu} \quad (8.1)$$

for some ‘matter’ energy-momentum tensor $T_{\mu\nu}$. The Lagrangian of a SUGRA model contains up to two (one) derivatives of bosonic (resp. fermionic) fields. The effective Lagrangian of a SUSY string theory is an *infinite* expansion in powers of momenta: the SUGRA Lagrangian L_{SUGRA} is just its truncation to the leading IR terms. We call this truncation the *supergravity approximation*.

Table 8.1 Spinors in d dimensions: M stands for Majorana, W for Weyl, MW for Majorana-Weyl, SM for symplectic-Majorana, and SMW for symplectic-Majorana-Weyl. The supercharges form N copies of basic spinor. In $d = 6, 10$, $N = p + q$ where p (q) counts spinors of chirality $+$ (resp. $-$)

d	10	9	8	7	6	5	4	3
Spinor	MW $\pm$	M	W	SM	SMW $\pm$	SM	W	M
UAut	$SO(p) \times SO(q)$	$SO(N)$	$U(N)$	$Sp(N)$	$Sp(p) \times Sp(q)$	$Sp(N)$	$U(N)$	$SO(N)$

There are several SUGRA models depending on the number d of spacetime dimensions, the specific SUSY symmetry (e.g. the number n_s of supercharges), the SUSY representation content of the ‘matter sector’, and other data.

In supergravity SUSY is a *gauged* symmetry, and we have one spin- $\frac{3}{2}$ gauge field $\psi_{\mu\alpha}^A$ —called a *gravitino*¹—per supercharge Q_α^A . $\psi_{\mu\alpha}^A$ transforms as a connection under a local SUSY transformation of parameter ϵ_α^A i.e.

$$\delta\psi_{\mu\alpha}^A = D_\mu\epsilon_\alpha^A + T_{\mu\alpha B}^{A\beta}\epsilon_\beta^B \quad (8.2)$$

for a suitable covariant derivative D_μ and a (known) *composite* Grassmann-even operator $T_{\mu\alpha B}^{A\beta}$. The R-symmetry group R ($\equiv$ the group of internal symmetries acting effectively on Q_α^A) is contained in the unitary automorphism group **UAut** of the SUSY algebra and contains its semisimple part **UAut**^{ss}.² The unitary automorphism groups of the diverse SUSY algebras in d dimensions are listed in Table 8.1. The table repeats in $d \bmod 8$ (Bott periodicity [8]) and is symmetric for $d \leftrightarrow 12 - d$.

In addition to the metric $g_{\mu\nu}$ and the gravitini $\psi_{\mu\alpha}^A$, a supergravity model may contain k -form gauge field of various degrees k (in particular gauge vectors A^m), scalar fields ϕ^i , and spin- $\frac{1}{2}$ fermions χ^a .

Local R-Symmetry Consistency requires the R-symmetry R to be also a local symmetry: indeed the SUSY gauge-field $\psi_{\mu\alpha}^A$ is charged under R and, in a consistent theory, a gauge-field may be charged only with respect to a gauge symmetry, as it happens in Yang-Mills theory. The derivative D_μ in (8.2) should be both spin- and R-symmetry covariant. The graviton supermultiplet contains *no* fundamental vector field which may play the role of a R-connection,³ so the R-connection R_μ^a should be a *composite operator* formed with the existing fields such as the scalars ϕ^i .

¹ μ, α are $Spin(d-1, 1)$ vector and spinor indices, respectively. A is the SUSY extension index on which the R-symmetry group acts.

² $R \neq \mathbf{UAut}$ only when the gravitational supermultiplet satisfies a reality constraint.

³ This follows from the representation theory of SUSY algebras [9]: it extends to the full connection in the covariant derivative D_μ the well-known fact that in GR the spin-connection ω_μ^{mn} is not a fundamental field but a composite operator of the vielbein e_μ^m and its derivatives.

We write $\mathcal{M}$ for the target-space where the scalars take value. The scalars' kinetic terms $\frac{1}{2}\sqrt{-g}G_{ij}(\phi)\partial^\mu\phi^j\partial_\mu\phi^i$ endow $\mathcal{M}$ with a Riemannian metric $G_{ij}(\phi)$. By general covariance, locally in field space the composite connection R_μ^a should have the general form

$$R_\mu^a = R^a(\phi)_i \partial_\mu\phi^i + \text{fermion bilinears} \quad (8.3)$$

for some coefficients $R(\phi)_i^a$ which may be functions of the scalars ϕ^i . We combine these functions into a 1-form $R^a \equiv R_i^a d\phi^i$ on $\mathcal{M}$ with coefficients in the Lie algebra $\mathfrak{r}$ of R . We conclude that the R-symmetry connection is the *pull-back* to spacetime of an $\mathfrak{r}$ -valued connection $R^a \equiv R_i^a d\phi^i$ on $\mathcal{M}$.

Geometry of the Scalar Manifold Compatibility of local R-symmetry with parallel transport in $\mathcal{M}$ requires its connection R^a to be part of the Levi-Civita connection of $\mathcal{M}$ i.e. the restricted holonomy group⁴ of $\mathcal{M}$ should have the form

$$\text{Hol}^0(\mathcal{M}) \subseteq R^{\text{ss}} \times C(R) \quad (8.4)$$

with $R^{\text{ss}} \equiv \mathbf{UAut}^{\text{ss}}$ the semi-simple part of the R-symmetry group and $C(R)$ its centralizer in $SO(\dim \mathcal{M})$. The R^{ss} -part of the Riemann curvature of $\mathcal{M}$ vanishes in rigid SUSY, since R is a *global* symmetry in that case, while it is covariantly-constant and negative in gauged SUSY i.e. SUGRA [1].

Note 8.1 We give a better argument which uses the notion of G -structure (see Chap. 11). The usual discussion of the SUSY algebra [10] and its representations [9] hold at the level of S -matrix: one has to choose a specific vacuum, which in the present case means a point $p \in \mathcal{M}$, and consider asymptotic **in/out** particle states described by small oscillations around the vacuum p . The fields describing asymptotic scalar particles in the vacuum p take value in the tangent space at p to the scalars' manifold, $T_p\mathcal{M}$. Hence the vector space $T_p\mathcal{M}$ decomposes into irreducible linear representations of the R-symmetry. These linear representations are determined by the supermultiplet content of L_{SUGRA} and so are independent of the chosen point p . The subspaces of $T_p\mathcal{M}$ associated with a given irreducible representation W of the R-symmetry then form a smooth bundle $\mathcal{V}_W$ over $\mathcal{M}$ and

$$T\mathcal{M} = \bigoplus_W \mathcal{V}_W. \quad (8.5)$$

By the Coleman-Mandula theorem [11], the R-symmetry acts unitarily on each $T_p\mathcal{M}$; hence the fibers of the sub-bundles $\mathcal{V}_W, \mathcal{V}_{W'}$, with $W' \neq W$ are orthogonal. Therefore the kinetic terms' metric G_{ij} endows the scalars' manifold $\mathcal{M}$ with a $\prod_W SO(\text{rank } \mathcal{V}_W)$ -structure which is torsion-less, and the decomposition (8.5) is preserved by parallel transport with the Levi-Civita connection of the metric G_{ij} .

⁴ We refer to Sect. 11.1.1 for background on Riemannian holonomy groups and the facts alluded in the text. In this Chapter we use well-known results from differential geometry in an informal fashion.

The Riemannian manifolds $\mathcal{M}$ which may appear as scalars' target-space of a SUSY theory are determined by the group embeddings

$$\begin{aligned} \mathrm{Hol}^0(\mathcal{M}) &\hookrightarrow C(R) \hookrightarrow SO(\dim \mathcal{M}) \quad (\text{rigid SUSY}) \\ \mathrm{Hol}^0(\mathcal{M}) &\hookrightarrow R^{\mathrm{ss}} \times C(R) \hookrightarrow SO(\dim \mathcal{M}) \quad (\text{SUGRA}) \end{aligned} \quad (8.6)$$

i.e. by the R-symmetry representation content of the theory [1]. This condition fixes all terms in the SUSY Lagrangian up to Yang-Mills and superpotential couplings. To illustrate the power of the method, we apply the rule to determine the well-known target-space geometries of 4d SUSY theories (rigid & local) with at most 8 supercharges:⁵

Examples in 4d with ≤ 8 supercharges

- (a) In 4d $\mathcal{N} = 1$ rigid SUSY $R = U(1)$ so that $\mathrm{Hol}^0(\mathcal{M}) = U(m)$ i.e. $\mathcal{M}$ is *Kähler* [15].
- (b) In 4d $\mathcal{N} = 1$ SUGRA the same story holds [16].
- (c) In 4d rigid SUSY $R = U(2)$. There are 2 types of matter multiplets, vector and hyper-multiplets, on which $SU(2)$ (resp. $U(1)$) acts trivially. Hence in 4d $\mathcal{N} = 2$ rigid SUGRA the vector multiplet (resp. hypermultiplet) scalars take value in a *Kähler* (resp. *hyperKähler*) manifold [17].
- (d) In 4d $\mathcal{N} = 2$ SUGRA the curvature of the $SU(2)$ factor is not zero hence the hypermultiplet scalars take value in a *negative quaternionic Kähler* manifold [18].

In 3d SUGRA with $2\mathcal{N}$ supercharges and $\mathcal{N} \geq 3$ we have (cf. Table 8.1)

$$\mathrm{Hol}^0(\mathcal{M}) \subseteq \mathrm{Spin}(\mathcal{N}) \times C(\mathrm{Spin}(\mathcal{N})). \quad (8.7)$$

The $\mathfrak{so}(\mathcal{N})$ part of the Riemannian curvature is covariantly-constant and negative [1, 19, 20]. By Berger's theorem [12, 21, 22] (to be reviewed in Sect. 11.1.1) for $\mathcal{N} \geq 4$ such a Riemannian manifold $\mathcal{M}$ is *locally symmetric of non-compact type*, i.e. of the form $\Gamma \backslash G(\mathbb{R})/K$ where $G(\mathbb{R})$ is a real Lie group, $K \subset G(\mathbb{R})$ a maximal compact subgroup fixed by a Cartan involution $\theta: G(\mathbb{R}) \rightarrow G(\mathbb{R})$, and $\Gamma \subset G(\mathbb{R})$ is a discrete subgroup, see [12, 23]. The group $G(\mathbb{R})$ is uniquely determined by the SUSY representation content i.e. how many supermultiplets of each kind. $G(\mathbb{R})$ determines the full Lagrangian up to the possibility of gauging a discrete subgroup

⁵ For the definition and basic properties of Kähler, hyperKähler, and quaternionic Kähler geometries see Sect. 11.1.1 or [12–14].

$\Gamma \subset G(\mathbb{R})$ or a continuous symmetry $\mathcal{G}$. The argument is in 3d, but we can always reduce to this case by toroidal compactification.⁶ Hence:

In any spacetime dimension $d \geq 3$ a SUGRA Lagrangian L_{SUGRA} with more than 8 supercharges is essentially determined by its SUSY representation content and the continuous internal gauge symmetry $\mathcal{G}$. The universal cover $\tilde{\mathcal{M}}$ of the scalars' target-space $\mathcal{M}$ is a symmetric space of non-compact type, so $\mathcal{M} = \Gamma \backslash G(\mathbb{R})/K$ for a discrete subgroup $\Gamma \subset G(\mathbb{R})$

'Essentially' means up to the discrete datum $\Gamma \subset C(\mathcal{G}) \subseteq G(\mathbb{R})$.

The Arithmetic Symmetry $G(\mathbb{Z})$

Before gauging a subgroup, the non-compact Lie group $G(\mathbb{R})$ is *naively* a symmetry of the theory. Such naive symmetries were dubbed “hidden symmetries” in the old SUGRA literature [24]. This naive continuous symmetry is spoiled by global conditions arising from Dirac quantization of gauge charges/fluxes [1]: *the actual symmetry* (before the gaugings) *is* $G(\mathbb{Z}) \subset G(\mathbb{R})$, *the arithmetic subgroup of* $G(\mathbb{R})$ *which leaves invariant the lattice of gauge charges/fluxes*. Then $\Gamma \subseteq G(\mathbb{Z})$, and the swampland conjectures require Γ to have finite index in $G(\mathbb{Z})$ [1]. The Γ -gauging breaks the $G(\mathbb{Z})$ symmetry down to $N(\Gamma)/\Gamma$, where $N(\Gamma) \subset G(\mathbb{Z})$ is the normalizer of Γ in $G(\mathbb{Z})$. The group $N(\Gamma)/\Gamma$ is at most finite and the strong version of the swampland conjectures require it to be trivial.

$\Gamma \backslash G(\mathbb{R})/K$ with $\Gamma \subset G(\mathbb{R})$ an *arithmetic* subgroup is precisely the structure of the spaces we found in Sect. 6.5 and 7.7 as moduli of toroidal compactifications. This is a remarkably deep fact. We gave an example of its consequences in Sect. 7.7.2.

Note 8.2 The maximal compact subgroup $K \equiv \text{Hol}^0(\mathcal{M}) \subset G(\mathbb{R})$ is simply related to the R-symmetry group. In particular, when our SUGRA with more than 8 supercharges is “pure”, i.e. contains only the gravity supermultiplet and no matter supermultiplet, one has $K \equiv R$. The last condition is automatically satisfied for $d \geq 4$ SUGRAs with more than 16 supercharges.

⁶ Toroidal compactification changes the scalar manifold $\mathcal{M}$ as Eq. (8.6) and Table 8.1 imply. However the scalar manifold of the higher dimensional theory is a totally geodesic subspace of the one in lower dimension, while all totally geodesic subspaces of a locally symmetric space is also locally symmetric. Hence the statement in the gray box holds in $d \geq 3$ dimensions if it holds in $d = 3$. We shall illustrate this point explicitly in Chap. 13 when studying group disintegration and U -duality.

8.2 Non-Renormalization Theorems. BPS Objects

The SUGRA approximation is quite powerful because there are several important physical observables which are given *exactly* by the supergravity answer. Thus the SUGRA truncation opens a panoramic window on the non-perturbative physics of the supersymmetric strings.

Quantum Non-Renormalization Theorems

When there are more than 8 supercharges,⁷ which is the case for all 10d SUSY strings, L_{SUGRA} is uniquely determined by the symmetries to the full non-linear level. In particular the dependence on the scalar fields ϕ^i of the couplings in L_{SUGRA} is fixed and cannot be further corrected by quantum effects: this implies a host of SUSY *non-renormalization theorems*. The string coupling is given by the v.e.v. of the dilaton $g = \exp \Phi$; therefore the exact non-linear dependence on the string coupling of the operators in the string quantum effective action containing at most two derivatives is fixed by SUSY to be the tree-level one, which means that they do not receive any quantum correction by string loop corrections (i.e. higher genus amplitudes) nor by non-perturbative effects. Thus, while the SUGRA truncation is only reliable when the momenta k and the space-time curvatures are parametrically small in string units, $\sqrt{\alpha'}k \ll 1$ and $\alpha'R_{\mu\nu\rho\sigma} \ll 1$, under these circumstances it remains valid for arbitrary high string coupling $g = \langle e^\Phi \rangle$, so it has a non-perturbative bearing. While the existence of such quantum non-renormalization theorems is an automatic consequence of the SUSY Ward identities with more than 8 supercharges, we can also prove them directly by computing higher loop corrections in string perturbation theory, see Chap. 10.

It remains to justify the claim that there are physically interesting observables for which the truncation to two derivatives is exact (we call them *SUSY protected* quantities). This is the goal of the next two paragraphs.

BPS Objects

An important class of SUSY protected quantities are the BPS objects. There are several kinds of them: BPS states, BPS field configurations, BPS solutions of the classical e.o.m., and BPS operators (point-like as well as extended). An object is BPS iff it is invariant for $\ell \geq 1$ linear combinations of the supercharges

$$Q^{(a)} \equiv c_A^{(a)\alpha} Q_\alpha^A, \quad a = 1, \dots, \ell, \quad (8.8)$$

where the $c_A^{(a)\alpha}$ are object-dependent numerical coefficients. A BPS object is automatically invariant under the bosonic symmetries generated by the charges $T^{(ab)} \equiv \{Q^{(a)}, Q^{(b)}\}$. If n_s is the total number of supercharges, an object invariant under ℓ of them is said to be $\frac{\ell}{n_s}$ -BPS. For instance, a state $|\text{BPS}\rangle$ is BPS iff $Q^{(a)}|\text{BPS}\rangle = 0$ and a bosonic operator Φ_{BPS} is BPS iff $[Q^{(a)}, \Phi_{\text{BPS}}] = 0$.

⁷ The border case of 8 supercharges will be considered in detail in Chap. 11.

We are particularly interested in BPS *classical field configurations*, which we take to be purely bosonic, i.e. all Fermi fields vanish. Such a configuration is BPS iff

$$\epsilon_{(a)} \delta^{(a)} \Phi \equiv \delta_{c_A^{(a)\alpha} \epsilon_{(a)}} \Phi = 0, \quad (8.9)$$

where the boldface symbol Φ stands for all fields in the theory, $\delta_{\epsilon_A^\alpha} \Phi$ is the classical SUSY variation of Grassmann parameter ϵ_A^α , while $\epsilon_{(a)}$ are the Grassmann-odd parameters of the *unbroken* SUSY. If the field Φ is bosonic, its variation is fermionic, hence automatically zero in our bosonic background. The only non-trivial BPS conditions arise from the variation of spin- $\frac{3}{2}$ and spin- $\frac{1}{2}$ fermions

$$\delta^{(a)} \psi_{\mu\alpha}^A = \delta^{(a)} \chi^m = 0. \quad (8.10)$$

Taking the Hermitian conjugate of these equations, we see that the set $Q^{(a)}$ of preserved supercharges is closed under Hermitian conjugation.

We are only interested in classical BPS configurations which are *on-shell*, i.e. solutions to the e.o.m. Such configurations correspond to backgrounds for string theory which preserve some SUSY: while we consider only the SUGRA truncation, BPS backgrounds are exact solution at the full string level by the argument in the next paragraph. The crucial point is that, for a classical configuration, the condition of being BPS is consistent with the equations of motion: SUSY invariance of the action yields

$$\sum_{\phi \text{ bos}} \frac{\delta S}{\delta \phi(x)} \delta^{(a)} \phi(x) + \sum_{\psi \text{ bos}} \frac{\delta S}{\delta \psi(x)} \delta^{(a)} \psi(x) = 0 \quad (8.11)$$

where ϕ (resp. ψ) stands for all the bosonic (resp. fermionic) fields in the action. Evaluated on our BPS bosonic background—but allowing for arbitrary Fermi fluctuations—the second term vanishes and we conclude that a large set of linear combinations of the bosonic e.o.m. $\delta S / \delta \phi(x) = 0$ are automatically satisfied. However, as it will be clear below, restricting the effective theory to the BPS configurations is *not* a consistent truncation.⁸ Depending on the situation, a BPS configuration may automatically satisfy all e.o.m. or we may need to impose a few more dynamical equations to put the full background on-shell.

For a suitable normalization of the constants $c_A^{(a)\alpha}$, and in the rest frame of our object, the SUSY sub-algebra generated by the $Q^{(a)}$'s has the schematic form

$$\{Q^{(a)\dagger}, Q^{(b)}\} = \delta^{ab} M - Z^{ab}, \quad (8.12)$$

where M is the mass, or more generally the tension (mass per unit volume), of the object, and Z^{ab} is a matrix of conserved charges (per unit volume). The RHS of (8.12) is non-negative: we get the BPS bound $M \geq \zeta$, where ζ is the largest eigenvalue of

⁸ A truncation of d.o.f. in a classical Lagrangian field theory is a *consistent truncation* iff all solutions of the truncated model are solutions to the original e.o.m. The truncation to configurations invariant under some *bosonic* symmetry is always consistent.

the matrix Z^{ab} evaluated on the given object. For a BPS object the LHS vanishes, the bound is saturated, and we have an equality for the charges $Z^{ab} = \delta^{ab} M$.

A BPS object belongs to a supermultiplet in which ℓ out of the n_s supercharges act trivially: the number of states in such a supermultiplet is $2^{[(n_s-\ell)/2]}$ instead of $2^{n_s/2}$ as for a generic supermultiplet. (A supermultiplet with $< 2^{n_s/2}$ states is called *short*). Therefore, if our object becomes BPS in some asymptotic limit—say at weak coupling $g \ll 1$, or in the low-momenta regime $k \ll 1/\sqrt{\alpha'}$ —it should remain BPS for all values of g and k because the number of states is continuous (hence constant) under continuous deformations of the parameters which preserve SUSY.⁹ This is an extension of the basic reasoning in Witten index theory [25]. In particular the BPS mass formula $M\delta^{ab} = Z^{ab}$ does not get renormalized. If, as we argue below, the operator Z^{ab} is not corrected, we get powerful non-renormalization theorems for the BPS sectors.

★ Global SUSY Charges in SUGRA

We consider configurations of our SUGRA which have the physical interpretation of excited states in a SUSY-preserving Poincaré invariant vacuum: the metric is asymptotic at infinity to the flat Minkowski one,¹⁰ while the SUSY variation $\delta_\epsilon \Psi$ of *all* fields Ψ (bosonic and fermionic) goes to zero as $r \rightarrow \infty$, i.e. the field configuration at infinity is a super-Poincaré invariant classical vacuum. In such a background the charges P_μ , Q_α^A , Z^{AB} of the super-Poincaré algebra are well defined and generate the usual SUSY algebra [26, 27] (see [1] for a review). In SUGRA the super-Poincaré generators are *gauge* charges which should obey a generalized Gauss' law, that is, each gauge charge q is given by the flux through the sphere at spatial infinity of a $(d-2)$ -form $\mathcal{E}_q$

$$q = \int_{S_\infty^{d-2}} \mathcal{E}_q, \quad \mathcal{E}_q \stackrel{\text{def}}{=} \frac{\partial(*L)}{\partial(\mathcal{F}_q)}, \quad (8.13)$$

where $*L$ is the SUGRA Lagrangian density and the $\mathcal{F}_q$ is the field-strength of the gauge connection ω_q associated to the symmetry generated by q .

For definiteness we focus on $d = 4$, the generalization to arbitrary d being straightforward. The only term in L containing derivatives of the gravitino is the Rarita-Schwinger kinetic term

$$L = \dots + \frac{i}{2} \bar{\psi}_\mu{}^A \gamma^{\mu\nu\rho} D_\nu \psi_\rho^A + \dots \quad (8.14)$$

The generalized Gauss' law yields an explicit expression for the supercharges as fluxes at infinity

$$\bar{\epsilon} Q \equiv \bar{\epsilon}_A^\alpha Q_\alpha^A = \frac{i}{2} \int_{S_\infty^2} \bar{\epsilon}_A \gamma_{\mu\nu\rho} \psi^{\rho A} dx^\mu \wedge dx^\nu \quad (8.15)$$

and (leaving all indices implicit)

$$[\bar{\epsilon}_1 Q, \bar{\epsilon}_2 Q] = i \delta_{\epsilon_1} \bar{\epsilon}_2 Q = \frac{1}{2} \int_{S^2} \bar{\epsilon}_2 \gamma_{\mu\nu\rho} \delta_{\epsilon_1} \psi^\rho dx^\mu \wedge dx^\nu. \quad (8.16)$$

⁹ **Warning.** There is one way the argument in the text may fail: several short supermultiplets can *recombine* together to form a *long* one. This phenomenon does not happen when the BPS supermultiplets preserve *many* supercharges (the supermultiplets are “*ultrashort*”) which is the case of interest in the present chapter.

¹⁰ More generally one can consider geometries which are asymptotic to anti-de Sitter (AdS), see [26, 27] or Chap. 6 of [1].

The insertion in (8.16) of the various terms in $\delta_{\epsilon_1} \psi_\rho$ (cf. Eq. (8.2)) produces the bosonic conserved charges which enters in $\{Q_\alpha^Q, Q_\beta^B\}$. The derivative term $D_\mu \epsilon_1$ yields $\bar{\epsilon}_1 \gamma^\mu \epsilon_2 P_\mu$ where P_μ is the Nester form [28, 29] of the conserved momentum in an asymptotic Minkowski space. The other terms yield conserved ‘central charges’ of various kinds. For instance, in a $d = 4$ SUGRA the gravitino transformation contains terms of the form

$$\delta_\epsilon \psi_\mu^A = D_\mu \epsilon_\mu^A + \dots - \frac{1}{4} \mathcal{F}_{\mu\nu}^{AB} \gamma^{\mu\nu} \gamma_\rho \epsilon_B + \text{fermi bilinears} \quad (8.17)$$

where $\mathcal{F}_{\mu\nu}^{AB}$ is a gauge-covariant, $G(\mathbb{R})$ -invariant, self-dual 2-form composite operator $\mathcal{F}_{\mu\nu}^{AB}$ in the 2-index antisymmetric representation of the R-group $SU(\mathcal{N})$. Plugged in the RHS of (8.16) this term produces a complex central charge Z^{AB} given by the flux at infinity of the self-dual 2-form

$$Z^{AB} = \int_{S_\infty^2} * \mathcal{F}^{AB}. \quad (8.18)$$

In the SUGRA truncation the composite 2-form operator must have the form

$$\mathcal{F}_{\mu\nu}^{AB} = f_a^{AB} F_{\mu\nu}^a + g_a^{AB} \tilde{F}_{\mu\nu}^a + \text{Fermi bilinears} \quad (8.19)$$

where $F_{\mu\nu}^a$ are the field-strengths of the gauge vectors and f_a^{AB}, g_a^{AB} are functions of the scalar fields (for their deep geometric meaning see Sects. 4.6 and 4.7 of [1]). The central charge Z^{AB} in (8.18) depends only on the asymptotic configuration at infinity. In particular only the value of the scalars at infinity matters for Z^{AB} ; these values correspond to a classical SUSY-preserving vacuum.

Regardless of all details, Z^{AB} , being a flux at infinity by Gauss’ law, depends only on the *large-scale* asymptotic behaviour of the fields: this is precisely the asymptotic limit of vanishing momenta where the SUGRA truncation becomes exact. The string coupling constant is given by the value of the dilaton at infinity $g = \exp[\Phi(\infty)]$, so the dependence of Z^{AB} on g is exactly given by the SUGRA approximation. For a BPS object the mass/tension is its central charge. We conclude that the dependence on g of the masses/tensions of BPS objects, as computed from SUGRA, is then non-perturbatively exact at the full string level (with the **Warning** in footnote 9).

Note 8.3 Since $\{Q_\alpha^A, Q_\beta^B\}$ is a non-negative operator equal to $\gamma^\mu P_\mu$ plus central charge, the above result entails that the mass is positive in supergravity. The above considerations (pushed to a more detailed level) imply the positive-mass theorem in General Relativity as observed by Witten [30] and others [26–29].

8.3 Supergravity in 11d

The largest possible space-time dimension in which one can have a locally Poincaré invariant SUSY theory is eleven. This arises from two considerations:

- WW** we cannot have a non-trivial theory which is *local* (in the sense that it has a conserved energy-momentum tensor $T_{\mu\nu}$) and contains *massless* particles of spin larger than 2. We may have at most *one* massless spin-2 field (the *graviton*). This is the content of the Weinberg-Witten theorem [31].¹¹
- SUSY** Fix the null momentum of the massless particles, and choose a plane transverse to it; define the helicity as the eigenvalue of the $U(1)$ rotation in

¹¹ The statement in the main text is slightly stronger than the Weinberg-Witten theorem.

this plane. Half of the n_s supercharges vanish on a state of null momentum. Half of the non-zero $n_s/2$ supercharges have helicity $-\frac{1}{2}$; we write

$$Q_-^a \quad (\text{with } a = 1, \dots, n_s/4) \quad (8.20)$$

for the non-zero helicity $-\frac{1}{2}$ supercharges. Let $|2\rangle$ be a state of the graviton with the given null momentum p and helicity 2: the states

$$|[a_1 \cdots a_s], 2 - s/2\rangle \stackrel{\text{def}}{=} Q_-^{a_1} \cdots Q_-^{a_s} |2\rangle, \quad 0 \leq s \leq n_s/4, \quad (8.21)$$

are non-zero, belong to the multiplet of the graviton, and have helicity $2 - s/2$. The state (8.21) with the lower helicity is obtained for $s = n_s/4$; by **WW** its helicity must be ≥ -2 , that is,

$$2 - n_s/8 \geq -2 \quad \text{i.e.} \quad n_s \leq 32. \quad (8.22)$$

Thus, any consistent theory should have $n_s \leq 32$ supercharges in any number of dimensions.¹² Since Q_α^A is a spinor, the maximal spacetime dimension d in which we have a consistent SUSY theory is the maximal dimension in which the number of real components of a spinor is less or equal to 32. The number of real components of the d -dimensional *minimal* spinors in Table 8.1 is given by the function $N(d)$ defined as¹³

$$\begin{cases} N(d+8) = 16 N(d) \\ N(3) = 2, \quad N(4) = 4, \quad N(5) = N(6) = 8, \\ N(7) = N(8) = N(9) = N(10) = 16. \end{cases} \quad (8.23)$$

The largest d for which $N(d) \leq 32$ is $d = 11$, where the bound is saturated by a single Majorana spinor.

11d SUGRA

Compactifying the 11d supergravity to 10d on a circle, we get a 10d SUGRA with 32 supercharges, in fact 16 supercharges of each chirality. Since the non-chiral supergravity with 32 supercharges in 10d is *unique* (cf. Sect. 8.1), this should also be the low-energy effective theory of the Type IIA superstring which is non-chiral with 32 supercharges. It is simpler to work in 11d than in 10d, because of the larger symmetry. In this chapter we use 11d supergravity as a mere technical *trick* to get the low-energy effective Lagrangian, but we shall see in Chap. 13 that this 11d-10d connection is much deeper than just a formal trick.

¹² Properly speaking, the argument in the text applies only for $d \geq 4$, since for $d \leq 3$ there is no propagating graviton i.e. no state $|2\rangle$. However the result remains true in 3d if we restrict to locally non-trivial theories, i.e. SUGRA models with propagating local d.o.f. [1].

¹³ By the standard relations between (universal) Clifford algebras in different signatures [32], the function defined here, $N(d)$, is equal to $2 \mathbf{N}(d-2)$ where $\mathbf{N}(k)$ is the function defined in Eq. (2.57) of [1] i.e. the real dimension of the irreducible modules of the Clifford $\mathbb{R}$ -algebra in dimension k .

To guess the field content of 11d SUGRA let us recall the massless spectrum of the Type IIA superstring (here $\mu, \nu = 0, 1, \dots, 9$ and $a = 1, 2$)

$$\begin{array}{ll} \text{NS-NS:} & G_{\mu\nu}, B_{\mu\nu}, \Phi \\ \text{NS-R, R-NS:} & \psi_\mu^a, \chi^a \\ \text{R-R:} & A_\mu, A_{\mu\nu\rho}, \end{array} \quad (8.24)$$

where all form-fields ($B_{\mu\nu}$, A_μ and $A_{\mu\nu\rho}$) are *gauge fields*. The fields $G_{\mu\nu}$, Φ and A_μ combine¹⁴ in the 11d metric G_{MN} ($M, N = 0, \dots, 10$). Likewise, ψ_μ^a and χ^a combine in a single 11d Majorana gravitino Ψ_M . The remaining degrees of freedom, $B_{\mu\nu}$ and $A_{\mu\nu\rho}$ combine in a 11d 3-form field¹⁵ $C_{MNP}^{(3)}$ such that

$$B_{\mu\nu} \sim C_{\mu\nu 10}^{(3)}, \quad A_{\mu\nu\rho} \sim C_{\mu\nu\rho}^{(3)}. \quad (8.25)$$

Equality in the numbers of fermionic and bosonic d.o.f. implies that the 3-form $C^{(3)}$ must be a gauge field with gauge transformation

$$\delta C^{(3)} = d\xi^{(2)}, \quad (8.26)$$

where ξ is an arbitrary 2-form. Its gauge invariant field strength is then

$$F^{(4)} = dC^{(3)}. \quad (8.27)$$

We note that in 11d we have no scalar, and hence the Lagrangian is polynomial in the “matter” fields Ψ_M , $C_{MNP}^{(3)}$ and their derivatives. Assuming no more than two derivatives, SUSY fixes the Lagrangian uniquely.

In this textbook we shall always write the supergravity Lagrangian up to four-fermions interactions, which—if needed—may be found in many places in the literature.¹⁶ Often we limit ourselves to the bosonic part of the Lagrangian. We normalize the 3-form field to match the conventions common in the string literature.

The Lagrangian is [33]

$$\begin{aligned} S = \frac{1}{2\kappa^2} \int d^{11}x \, e \left[e^{AM} e^{BN} R_{MNAB} - \bar{\Psi}_M \Gamma^{MNP} D_N \Psi_P - \frac{1}{2} |F^{(4)}|^2 - \right. \\ \left. - \frac{1}{96} \bar{\Psi}_M (\Gamma^{IJKLMN} + 12 \Gamma^{IJ} G^{MK} G^{NL}) \Psi_\rho F_{IJKL}^{(4)} \right] - \\ - \frac{1}{12\kappa^2} \int C^{(3)} \wedge F^{(4)} \wedge F^{(4)} + \text{4-Fermi interactions}, \end{aligned} \quad (8.28)$$

¹⁴ Here ‘combine’ means identification at the linearized level (which is enough to identify and count d.o.f.). The full non-linear relation between 11d and 10d fields is, of course, more involved.

¹⁵ Here and below a symbol of the form $X^{(k)}$ means that the field X is a k -form.

¹⁶ See, for instance, for 11 SUGRA in [33], for 10d IIA SUGRA in [34, 35], for 10d IIB SUGRA in [36], and for 10d $\mathcal{N} = 1$ SUGRA with matter in [37].

where e_M^A is the 11-bein, $G_{MN} \equiv e_M^A e_N^B \eta_{AB}$, and $e = \det e_M^A = \sqrt{-G}$. The norm of a p -form is

$$|X^{(p)}|^2 \stackrel{\text{def}}{=} \frac{1}{p!} X^{M_1 \dots M_p} X_{M_1 \dots M_p}. \quad (8.29)$$

The above action is invariant under the following SUSY transformations [33]:

$$\delta e_M^A = \frac{1}{2} \bar{\epsilon} \Gamma^A \Psi_M, \quad (8.30)$$

$$\delta \Psi_M = D_M \epsilon + \frac{1}{288} (\Gamma^{IJKL}{}_M - 8 \Gamma^{JKL} \delta_M^I) F_{IJKL}^{(4)} \epsilon + 3\text{-fermions} \quad (8.31)$$

$$\delta C_{MNP}^{(3)} = -\frac{3}{2} \bar{\epsilon} \Gamma_{[MN} \Psi_{P]}. \quad (8.32)$$

8.4 Type IIA Superstring: Low-Energy Effective Theory

Next we get 10d Type IIA *supergravity* by compactifying 11d SUGRA on a circle S^1 . 10d IIA SUGRA was first constructed in [34, 35, 38].

Type IIA Supergravity

We compactify 11d SUGRA down to 10d on a circle of length $2\pi R$. We know from Sect. 6.1 that all 11d metric invariant by translations along the circle have the form¹⁷

$$ds^2 = G_{MN}^{(11)} dx^M dx^N = G_{\mu\nu}(x^\rho) dx^\mu dx^\nu + e^{2\sigma(x^\mu)} [dx^{10} + A_\nu(x^\rho) dx^\nu]^2 \quad (8.33)$$

where $M, N = 0, \dots, 10$ and $\mu, \nu = 0, \dots, 9$. The 11 metric $G_{MN}^{(11)}$ reduces to the 10d metric $G_{\mu\nu}$ together with a 10d vector A_ν , which is a gauge field because of the invariance under the diffeomorphism

$$A_\nu(x^\rho) \rightarrow A_\nu(x^\rho) + \partial_\mu \lambda(x^\rho), \quad x^{10} \rightarrow x^{10} - \lambda(x^\rho), \quad (8.34)$$

and a scalar σ . Likewise the 3-form $C^{(3)}$ reduces to

$$B_{\mu\nu} = C_{\mu\nu 10}^{(3)} \quad \text{and} \quad A_{\mu\nu\rho} = C_{\mu\nu\rho}^{(3)}. \quad (8.35)$$

We write $A^{(1)}$, $A^{(2)}$, $A^{(3)}$ for, respectively, the 10d form-fields A_μ , $B_{\mu\nu}$, $A_{\mu\nu\rho}$, and $F^{(2)} = dA^{(1)}$, $F^{(3)} = dA^{(2)}$, $F^{(4)} = dA^{(3)}$ for their gauge-invariant field strengths; the superscript is the degree as differential forms.

¹⁷ For Kaluza-Klein geometries in supergravity see the review [39].

We focus on the bosonic part of the action (8.28) which we divide in three terms

$$S_1 = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G} \left(e^\sigma R - \frac{1}{2} e^{3\sigma} |F^{(2)}|^2 \right) \quad (8.36)$$

$$S_2 = -\frac{1}{4\kappa_{10}^2} \int d^{10}x \sqrt{-G} \left(e^{-\sigma} |F^{(3)}|^2 + e^\sigma |\tilde{F}^{(4)}|^2 \right) \quad (8.37)$$

$$S_3 = -\frac{1}{4\kappa_{10}^2} \int A^{(2)} \wedge F^{(4)} \wedge F^{(4)} = -\frac{1}{4\kappa_{10}^2} \int A^{(3)} \wedge F^{(3)} \wedge F^{(4)}. \quad (8.38)$$

Here $\kappa_{10}^2 = \kappa^2/2\pi R$ and we have defined

$$\tilde{F}^{(4)} = dA^{(3)} - A^{(1)} \wedge F^{(3)} \quad (8.39)$$

which is the improved gauge-invariant field strength (see **BOX 8.1** for a proof).¹⁸

For a proof of Eqs. (8.36)–(8.38), see **BOX 8.2**. For the full dimensional reduction, including fermions and 4-fermion terms, see the original literature [34, 35, 38].

Exercise 8.1 Write the SUSY variations of the fields using 11d $\rightarrow$ 10d reduction.

Note 8.4 The action contains many terms where p -form potentials $A^{(p)}$ appear directly rather than through their exterior derivatives $F^{(p+1)} = dA^{(p)}$, but the action is still gauge invariant (up to a surface term). Such couplings are known as *Chern-Simons terms*, and there are of two kinds. One involves the wedge product of a potential with any number of field strengths and it is gauge invariant as a consequence of the Bianchi identity for field strengths $dF^{(p+1)} = 0$. The other type appears in the kinetic term for the improved field strength (8.39). As discussed in **BOX 8.1**, the second term in $\tilde{F}^{(4)}$ has a gauge transformation

$$-d\lambda \wedge F^{(3)} = -d(\lambda F^{(3)}), \quad (8.40)$$

which is cancelled by the transformation

$$\delta A^{(3)} = \lambda F^{(3)}, \quad (8.41)$$

which is inherited from 11d **Diff**-invariance. $\tilde{F}^{(4)}$ is then invariant under all gauge transformations, but has the modified Bianchi identity

$$d\tilde{F}^{(4)} = -F^{(2)} \wedge F^{(3)}. \quad (8.42)$$

Note that the 0-form gauge transformation of S_3 vanishes since

$$\delta(A^{(3)} \wedge F^{(4)} \wedge F^{(3)}) = \lambda F^{(3)} \wedge F^{(4)} \wedge F^{(3)} + A^{(3)} \wedge d\lambda \wedge F^{(3)} \wedge F^{(3)} \equiv 0. \quad (8.43)$$

¹⁸ *Mutatis mutandis* this is the same mechanism as discussed in **BOX 6.2**.

BOX 8.1 - The improved field strength $\tilde{F}^{(4)}$

In 11d $dC^{(3)}$ is covariant under all diffeomorphisms, including the ones (8.34) which become the $A^{(1)}$ gauge symmetry in 10d. Under that diffeomorphism

$$\delta(dC^{(3)})_{\mu\nu\rho\sigma} = 4(dC^{(3)})_{[\mu\nu\rho 10} \partial_{\sigma]}\lambda$$

or, in the 10d notation,

$$\delta F^{(4)} = \delta A^{(1)} \wedge F^{(3)}$$

with $\delta A^{(1)} = d\lambda$. The gauge invariant combination is $\tilde{F}^{(4)} = F^{(4)} - A^{(1)} \wedge F^{(3)}$; this is the physically well-defined quantity, invariant under *all* (zero-and two-form) gauge transformations. It satisfies the peculiar *Bianchi identity*

$$d\tilde{F}^{(4)} = -F^{(2)} \wedge F^{(3)}.$$

Type IIA Superstring at Low-Energy

By uniqueness (Sect. 8.1) the fields of IIA SUGRA must correspond to the massless degrees of freedom of the IIA superstring. In particular the scalar σ must be the dilaton Φ , up to some field redefinition. The terms in the action are proportional to $e^{-k\sigma}$ for various values of k . The string coupling constant is determined by the value of the dilaton; after the appropriate field redefinitions, the tree-level spacetime Type IIA action should be multiplied by an overall factor $e^{-2\Phi}$ (because the sphere has $\chi = 2$) and otherwise depends on Φ only through its derivatives. Here ‘appropriate field redefinitions’ means that the “new” fields must be the ones entering in the string world-sheet action as the natural couplings (the *string frame* fields cf. Sect. 1.8).

Our next task is to find the field redefinition mapping the fields of the standard Lagrangian formulation of IIA SUGRA (*Einstein frame*) into the ones which appear in the IIA world-sheet action. We write the expression of the new fields in terms of the old ones; these formulae will be justified in the next paragraph. First of all we redefine the fields

$$G_{\mu\nu} \Big|_{\text{old}} = e^{-\sigma} G_{\mu\nu} \Big|_{\text{new}}, \quad \sigma = \frac{2\Phi}{3}. \quad (8.44)$$

From now on $G_{\mu\nu}$ always stands for the new metric in the *string frame*. Using the formulae in BOX 1.12 with $d = 10$ and $f = -\sigma/2$, we get

$$\begin{aligned} \sqrt{-G} e^{\sigma} R &\rightarrow \sqrt{-G} e^{-3\sigma} \left(R + 9 D^{\mu} \partial_{\mu} \sigma - 18 \partial_{\mu} \sigma \partial^{\mu} \sigma \right) = \\ &= \sqrt{-G} e^{-3\sigma} \left(R + 9 \partial_{\mu} \sigma \partial^{\mu} \sigma \right) + \text{surface term} = \\ &= \sqrt{-G} e^{-2\Phi} \left(R + 4 \partial_{\mu} \phi \partial^{\mu} \phi \right) + \text{surface term}. \end{aligned} \quad (8.45)$$

BOX 8.2 - Proofs of Eqs. (8.36) and (8.37)

Proof of Eq. (8.36) The (inverse) 11-bein corresponding to the metric (8.33) is

$$E_M{}^A = \left(\frac{e_\mu{}^a}{0} \middle| \frac{e^\sigma A_\mu}{e^\sigma} \right), \quad E_A{}^N = \left(\frac{e_a{}^v}{0} \middle| \frac{-A_a}{e^{-\sigma}} \right), \quad (\clubsuit)$$

so that

$$\sqrt{-G_{(11)}} = \det E = e^\sigma \det e = e^\sigma \sqrt{-G}.$$

By definition, the 11d Ricci scalar $R_{(11)}$ is^a

$$\begin{aligned} -4R_{(11)} &= \Omega_{ABC} \Omega^{ABC} - 2\Omega_{ABC} \Omega^{CAB} - 4\Omega_{CA}{}^A \Omega^C{}_B{}^B \\ \text{where } \Omega_{ABC} &= 2E_A{}^M E_B{}^N \partial_{[N} E_{M]C}, \end{aligned}$$

while the 10d Ricci scalar R is given by the same equations with $E_M{}^A$ replaced by the 10-bein $e_\mu{}^a$. The non-zero Ω_{ABC} are

$$\Omega_{abc} = 2e_a{}^\mu e_b{}^\nu \partial_{[\nu} e_{\mu]c}, \quad \Omega_{ab10} = -e^\sigma F_{ab}^{(2)}, \quad \Omega_{a1010} = -\Omega_{10a10} = -\partial_a \sigma.$$

Thus

$$\begin{aligned} -4R_{(11)} &= -4R + 2e^{2\sigma} |F^{(2)}|^2 \\ \sqrt{-G_{(11)}} R_{(11)} &= \sqrt{-G} \left(e^\sigma R - \frac{1}{2} e^{3\sigma} |F^{(2)}|^2 \right) + \text{surface term.} \end{aligned}$$

Finally,

$$\int d^{11}x (\dots) = 2\pi i R \int d^{10}x (\dots)$$

Proof of Eq. (8.37) From eq. (♣) one has

$$F_{abcd} \Big|_{11d} = F_{abcd}^{(4)} - 4A_{[a}^{(1)} F_{bcd]}^{(3)} \equiv \tilde{F}_{abcd}^{(4)}, \quad F_{abc10} \Big|_{11d} = e^{-\sigma} F_{abc}^{(3)}$$

^a This formula holds up to a term which when multiplied by $\sqrt{-G_{(11)}}$ becomes an irrelevant total derivative

Using this relation together with the transformation of the norms of differential p -forms under the Weyl rescaling $G_{\mu\nu} \rightarrow e^{2f} G_{\mu\nu}$ i.e.

$$\sqrt{-G} |F^{(p)}|^2 \rightarrow e^{(d-2p)f} \sqrt{-G} |F^{(p)}|^2, \quad (8.46)$$

we rewrite the IIA action in the form

$$S_{\text{IIA}} = S_{\text{NS}} + S_{\text{R}} + S_{\text{CS}}, \quad (8.47)$$

where

$$S_{\text{NS}} = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G} e^{-2\Phi} \left(R + 4 \partial^\mu \Phi \partial_\mu \Phi - \frac{1}{2} |H^{(3)}|^2 \right) \quad (8.48)$$

$$S_{\text{R}} = -\frac{1}{4\kappa_{10}^2} \int d^{10}x \sqrt{-G} \left(|F^{(2)}|^2 + |\tilde{F}^{(4)}|^2 \right), \quad (8.49)$$

$$S_{\text{CS}} = -\frac{1}{4\kappa_{10}^2} \int B^{(2)} \wedge F^{(4)} \wedge F^{(4)}, \quad (8.50)$$

and we changed the name of the fields $A^{(2)} \rightarrow B^{(2)}$ and $F^{(3)} \rightarrow H^{(3)} \equiv dA^{(2)}$ to agree with the traditional notation in string theory. In Eq. (8.47) we have separated the various terms according to whether the fields belong to the NS-NS of R-R sectors of the IIA superstring: the Chern-Simons term contains fields from both sectors. It is useful to single out fields from different sectors, so we adopt the following notation:

- for R-R potentials and field strengths $C^{(p)}$ and $F^{(p+1)}$;
- for NS-NS form field and field strength $B^{(2)}$ and $H^{(3)}$;
- for Type I and heterotic we use $A^{(1)}$ and $F^{(2)}$ for the 10d non-Abelian gauge fields.

The NS part of the action now depends on the dilaton through the standard overall factor $e^{-2\Phi}$. The R action does not have the expected factor $e^{-2\Phi}$, but it can be brought to this form by the further redefinition

$$C^{(1)} = e^{-\Phi} C^{(1)'}, \quad C^{(3)} = e^{-\Phi} C^{(3)'} \quad (8.51)$$

after which

$$\int d^{10}x \sqrt{-G} |F^{(2)}|^2 = \int d^{10}x \sqrt{-G} e^{-2\Phi} |F^{(2)'}|^2 \quad (8.52)$$

$$F^{(2)'} = dC^{(1)'} - d\Phi \wedge C^{(1)'}, \quad (8.53)$$

and similarly for $C^{(3)}$ and $F^{(4)}$. This redefinition makes explicit the dependence on the dilaton Φ , but at the cost of cluttering up the Bianchi identity and the gauge transformation which become

$$dF^{(2)'} = d\Phi \wedge F^{(2)'}, \quad \delta C^{(1)'} = d\lambda' - \lambda' d\Phi. \quad (8.54)$$

For this reason, the simpler Eqs. (8.48)–(8.50) are usually used.

Exercise 8.2 Show that in a background with a time-dependent dilaton, the conserved charge couples to the *unprimed* form-field.

Stringy Interpretation: BRST Cohomology Let us now make contact with string theory, and see why the background R-R fields appearing in the world-sheet action have the more complicated properties (8.54) while the NS-NS field $B^{(2)}$ behaves simply. We work at the linearized level in terms of the $(-\frac{1}{2}, -\frac{1}{2})$ R-R vertex operator

$$f_{\mu_1 \dots \mu_p}^{(p)}(X) S_\alpha(\Gamma^{\mu_1 \dots \mu_p})^{\alpha\dot{\beta}} \tilde{S}_{\dot{\beta}} e^{-\phi/2 - \tilde{\phi}/2}, \quad (8.55)$$

As seen in Sect. 3.7.3, this vertex is BRST-invariant iff

$$df^{(p)} = d * f^{(p)} = 0. \quad (8.56)$$

Equation (8.56) have the same form as the field equations and Bianchi identities for a p -form field-strength. In facts, as discussed in Sect. 3.7.3, $f^{(p)}(X)$ should be identified with the R-R *field-strength* rather than the potential. Note that we get only field-strengths of *even* degree, as we should for Type IIA. As already emphasized, R-R amplitudes will always contain a power of momentum, and so vanish at zero momentum.¹⁹

The BRST conditions (8.56) were obtained as the linearized equations of motion around a *flat* background. Now let us consider the effect of a dilaton gradient $d\Phi \neq 0$. When $\partial_\mu \Phi$ is a non-zero *constant* we get the linear dilaton background which, from the world-sheet perspective, is still a *free* CFT, cf. Sect. 2.5.5 and Sect. 1.8. In this CFT:

$$T_F = i(2/\alpha')^{1/2} \psi^\mu \partial X_\mu - 2i(\alpha'/2)^{1/2} \partial_\mu \Phi \psi^\mu, \quad (8.57)$$

$$G_0 = (\alpha'/2) \psi_0^\mu (p_\mu + i \partial_\mu \Phi) + \text{oscillators}, \quad (8.58)$$

and similarly for $\tilde{T}_F, \tilde{G}_0$. The net effect on the the BRST condition is to shift the spacetime derivatives by the dilaton gradient $\partial_\mu \rightarrow \partial_\mu - \partial_\mu \Phi$; Eq. (8.56) become

$$(d - d\Phi \wedge) f^{(p)} = 0, \quad (d - d\Phi \wedge) * f^{(p)} = 0. \quad (8.59)$$

The Bianchi identities and the field equations are modified precisely in the fashion we deduced from SUSY invariance of the effective space-time action at the full non-linear level, see Eq. (8.54). There is no such modification in the NS-NS tensor. $B^{(2)}$ couples to the world-sheet Σ directly through its potential

$$\frac{1}{2\pi\alpha'} \int_\Sigma B^{(2)}. \quad (8.60)$$

This coupling is invariant under $\delta B^{(2)} = d\xi^{(1)}$ independently of the dilaton, so $H^{(3)} = dB^{(2)}$ is invariant and $dH^{(3)} = 0$. As emphasized in the Chap. 6, perturbative winding states are electrically charged under this gauge symmetry. Put differently: the $(-1, -1)$ -picture B -vertex $\psi_{[\mu} \tilde{\psi}_{\nu]} e^{-\phi - \tilde{\phi}} e^{ikX}$ is BRST-invariant in *all* superconformal backgrounds since it is the first component of a conformal superfield.

Quasi-Topological Modes: Massive IIA Supergravity

The zero-momentum subtleties of the BRST cohomology (cf. Footnote 19) has an important consequence. The Type IIA string has quasi-topological modes which are not part of IIA SUGRA. Switching on these modes leads to a generalization of IIA SUGRA [40] which has no simple 11d origin but plays a role in string theory.

¹⁹ But recall the *crucial subtleties* with the BRST invariance at zero momentum—see the final part of Sect. 3.7.3, BOX 3.3, and Sect. 5.6.

We know from the Kähler-Dirac analysis of the BRST cohomology in the R-R sector (cf. **BOX 3.3**) that Type IIA contains a 10-form field-strength

$$F^{(10)} = dC^{(9)}. \quad (8.61)$$

The linearized Kähler-Dirac equation reduces to

$$d * F^{(10)} = 0, \quad (8.62)$$

and since $*F^{(10)}$ is a scalar, this means that $*F^{(10)}$ is frozen at zero momentum, so it is a non-propagating quasi-topological d.o.f. Nevertheless such a field has a physical effect, since it carries an energy density: a background with $*F^{(10)} \neq 0$ in 10d is analogous to a background electric field $*F^{(2)} \neq 0$ in 2d, where the electromagnetic field does not describe propagating photons but yields an electrostatic energy density associated to a linear Coulomb potential which confines electric charges [41].

It follows that the 9-form gauge field $C^{(9)}$ can be included in IIA supergravity. The resulting bosonic action is [40]

$$S'_{\text{IIA}} = \tilde{S}_{\text{IIA}} - \frac{1}{4\kappa_{10}^2} \int d^{10}x \sqrt{-G} M^2 + \frac{1}{2\kappa_{10}^2} \int M F^{(10)} \quad (8.63)$$

Here $\tilde{S}_{\text{IIA}}$ is the action in Eqs. (8.48)–(8.50) with the replacements

$$F^{(2)} \rightarrow F^{(2)} + MB^{(2)}, \quad (8.64)$$

$$F^{(4)} \rightarrow F^{(4)} + \frac{1}{2M} (F^{(2)} + MB^{(2)}) \wedge (F^{(2)} + MB^{(2)}), \quad (8.65)$$

$$\tilde{F}^{(4)} \rightarrow \tilde{F}^{(4)} + \frac{1}{2M} (F^{(2)} + MB^{(2)}) \wedge (F^{(2)} + MB^{(2)}). \quad (8.66)$$

Note that when $M \neq 0$ the vector's field-strength $F^{(2)}$ can be absorbed in $B^{(2)}$ by a gauge transformation

$$B^{(2)} \rightarrow B^{(2)} - \frac{1}{M} dA^{(1)}, \quad (8.67)$$

and the vector $A^{(1)}$ gets *eaten up* by the 2-form gauge field $B^{(2)}$ which becomes massive (with mass M). This is the 2-form version of the *Higgs mechanism*.

The $C^{(9)}$ e.o.m. say that on-shell M is a constant, hence a quasi-topological d.o.f. frozen at zero momentum. The scalar M appears in the action without derivatives and only quadratically. M can be easily integrated out, leading to a messy non-linear dependence on $B^{(2)}$. However this is *not* the proper procedure for a quasi-topological mode as M : physically M is just an integration constant in the equations of motion which is fixed by the boundary conditions.

For further details on massive IIA SUGRA see [42] and references therein.

Note 8.5 The deep physical reason why massive IIA SUGRA has no 11-dimensional uplift will be explained in Sect. 13.8.

8.5 Type IIB: Effective Low-Energy Theory

We know from the gray box in Sect. 8.1 that the low-energy theory of Type IIB, which is a SUGRA with two MW gravitini of the *same* chirality, is uniquely determined by supersymmetry up to the possible gauging of a discrete symmetry Γ . We consider first the ‘basic’ IIB SUGRA in which no discrete symmetry is gauged.²⁰

Naive Hidden Symmetry

Neglecting for the moment flux quantization, the ‘basic’ IIB model has a $SL(2, \mathbb{R})$ non-compact “hidden” global symmetry. Indeed IIB SUGRA has a $SO(2) \simeq U(1)$ R-symmetry which rotates the two MW gravitini. The model has 32 supercharges, so is “pure” in the sense of Note 8.2. Comparing with Sect. 8.1 we conclude that the covering scalar space $\tilde{\mathcal{M}}$ is a negatively-curved symmetric space of holonomy $U(1)$. There is precisely *one* such symmetric space, the upper half-plane $\mathcal{H} \equiv SL(2, \mathbb{R})/U(1)$ with isometry group $SL(2, \mathbb{R})/\{\pm 1\}$. Thus $SL(2, \mathbb{R})$ is a naive symmetry of the ‘basic’ model. The actual symmetry is the arithmetic subgroup consistent with flux quantization, namely $SL(2, \mathbb{Z})$. All other IIB SUGRAs are obtained by gauging a discrete²¹ subgroup $\Gamma \subset SL(2, \mathbb{Z})$. The R-symmetry in particular predicts the existence of two real massless scalars in Type IIB which take value in $\mathcal{H}$.

Equations of Motion

Type IIB supergravity has *no* Lagrangian in the conventional sense. This is due to the presence of the self-dual field strength $\tilde{F}^{(5)} = *\tilde{F}^{(5)}$ whose 4-form gauge potential $A^{(4)}$ is a chiral boson. There is no simple covariant action for such a field. However for the mere purpose of writing the classical equations of motion and checking their symmetries, we can work with the following action [45, 46] subjected to the limitations to be specified momentarily. We write only the bosonic terms:

$$S_{\text{IIB}} = S_{\text{NS}} + S_{\text{R}} + S_{\text{CS}} \quad (8.68)$$

$$S_{\text{NS}} = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G} e^{-2\phi} \left(R + 4 \partial^\mu \phi \partial_\mu \phi - \frac{1}{2} |H^{(3)}|^2 \right) \quad (8.69)$$

$$S_{\text{R}} = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G} \left(|F^{(1)}|^2 + |\tilde{F}^{(3)}|^2 + \frac{1}{2} |\tilde{F}^{(5)}|^2 \right) \quad (8.70)$$

$$S_{\text{CS}} = -\frac{1}{4\kappa_{10}^2} \int C^{(4)} \wedge H^{(3)} \wedge F^{(3)}, \quad (8.71)$$

²⁰ 10d IIB SUGRA was first constructed in [43, 44].

²¹ We cannot gauge a continuous subgroup since there are no vector fields.

BOX 8.3 - ‘Basic’ IIB SUGRA: further details

The $SO(8) \times U(1)_R$ representations in the *linearized* graviton supermultiplet is ([9] table 1)

$$[\mathbf{1}_{-2} \oplus (\mathbf{28}_v)_{-1} \oplus (\mathbf{35}_v)_0 \oplus (\mathbf{35}_-)_{-1} \oplus (\mathbf{28}_v)_2 \oplus \mathbf{1}_{-2}] \oplus [(\mathbf{8}_+)_{-\frac{3}{2}} \oplus (\mathbf{56}_+)_{-\frac{1}{2}} \oplus (\mathbf{56}_+)_{\frac{1}{2}} \oplus (\mathbf{8}_+)_{\frac{3}{2}}]$$

where the first (second) bracket contains bosonic (fermionic) states. At the full-non-linear level, the *Einstein frame* bosonic fields transform under the *global* $SL(2, \mathbb{R})$ symmetry and the fermions under the *local* $U(1)_R$. The gravitinos have $U(1)_R$ charges $\pm \frac{1}{2}$ and the dilatinos $\pm \frac{3}{2}$. The metric and the 4-form gauge-field are $SL(2, \mathbb{R})$ invariant while the two 3-forms $F_i^{(3)}$ field-strengths transform as a $SL(2, \mathbb{R})$ doublet. Let $E(x) \in SL(2, \mathbb{R})$ be a representative of the scalar configuration $\tau(x) \in \mathcal{H}$; $E(x)$, $E(x)'$ represent the same scalar configuration iff $E(x)' = E(x)U(x)$ with $U(x) \in U(1)_R$. By general theory of symmetric spaces [23] (cf. Chap. 5 of [1] or BOX 6.7) the $U(1)_R$ composite connection Q_μ and the target metric vielbein P_μ^a ($a = 1, 2$) (i.e. $G(\phi)_{ij} \partial^\mu \phi^i \partial_\mu \phi^j \equiv P_\mu^a P_\mu^a$) are obtained by decomposing the Maurier-Cartan form into eigenspaces of the Cartan involution θ i.e.

$$E^{-1} \partial_\mu E = \sigma_{2a-1} P_\mu^a + i \sigma_2 Q_\mu$$

(the σ_i are Pauli matrices). The covariant derivatives on the gravitino and dilatino are then

$$D_\mu \psi_\nu = (\nabla_\mu - \frac{i}{2} Q_\mu) \psi_\nu, \quad D_\mu \lambda = (\nabla_\mu - \frac{3i}{2} Q_\mu) \lambda$$

where ∇_μ is the covariant derivative for the spin-connection. One defines^a $SL(2, \mathbb{R})$ -invariant 3-form field-strengths of definite $U(1)_R$ charge

$$G^\pm = \frac{1}{2} (1 \pm \sigma_2) E^t \Omega \begin{pmatrix} F_1^{(3)} \\ F_2^{(3)} \end{pmatrix}$$

The SUSY variations of fermions then read (setting $G^+ = G$, $G^- = G^*$)

$$\begin{aligned} \delta \lambda &= i \gamma^\mu \epsilon^* P_\mu - \frac{i}{24} \gamma^{\mu\nu\rho} \epsilon G_{\mu\nu\rho} + \dots \\ \delta \psi_\mu &= D_\mu \epsilon + \frac{i}{480} \gamma^{\rho_1 \dots \rho_5} \gamma_\mu \epsilon \tilde{F}_{\rho_1 \dots \rho_5} + \frac{1}{96} (\gamma_\mu^{\nu\rho\sigma} G_{\nu\rho\sigma} - 9 \gamma^{\rho\sigma} G_{\mu\rho\sigma}) \epsilon^* + \dots \end{aligned}$$

where $D_\mu \epsilon = (\partial_\mu + \frac{1}{4} \omega_\mu^{rs} \gamma_{rs} - \frac{i}{2} Q_\mu) \epsilon$ and $\dots$ stands for terms quadratic in the fermions. In the Iwasawa gauge^b we have ($\tau = \tau_1 + i \tau_2 \in \mathcal{H}$)

$$E = \begin{pmatrix} \tau_2^{1/2} & \tau_1 \tau_2^{-1/2} \\ 0 & \tau_2^{-1/2} \end{pmatrix}, \quad G = i \tau_2^{-1/2} (F_1 + \tau F_2)$$

^a For the theory of $G(\mathbb{R})$ -invariant field-strengths in extended SUGRA see Sect. 4.7.1 of [1]

^b For the theory of the Iwasawa gauge in SUGRA see Sect. 5.11 of [1]

where, as always, tilded symbols stand for *improved* field-strengths:

$$\tilde{F}^{(3)} = F^{(3)} - C^{(0)} \wedge H^{(3)} \quad (8.72)$$

$$\tilde{F}^{(5)} = F^{(5)} - \frac{1}{2} C^{(2)} \wedge H^{(3)} + \frac{1}{2} B^{(2)} \wedge F^{(3)}. \quad (8.73)$$

The couplings between NS-NS fields are (of course) the same ones as in Type IIA. The e.o.m. and Bianchi identities for $\tilde{F}^{(5)}$ which follow from the action (8.68) are

$$d * \tilde{F}^{(5)} = d\tilde{F}^{(5)} = H^{(3)} \wedge F^{(3)}. \quad (8.74)$$

The light spectrum of the IIB superstring includes a *self-dual* 5-form field-strength. The field equations from the action (8.68) are consistent with the constraint

$$* \tilde{F}^{(5)} = \tilde{F}^{(5)} \quad (8.75)$$

but they do not imply it. To get the right theory we need to impose (8.75) as an *ad hoc* constraint on the solutions. It cannot be imposed directly on the action since $|\tilde{F}^{(5)}|^2 \equiv 0$ for a self-dual 5-form. Imposing the constraint is essential for SUSY since it ensures that there is an equal number of bosonic and fermionic d.o.f.

This formulation with an *ad hoc* condition on the solutions is (barely) acceptable in a classical treatment, but the self-duality constraint becomes very subtle in the quantum theory. We use the above action *only* for classical considerations, and refer the reader to the literature for a proper quantum treatment: see e.g. [47].

The $SL(2, \mathbb{R})$ Symmetry Unveiled

We consider the ‘basic’ IIB SUGRA with no discrete gauging. We work with the space-time metric in the Einstein frame (as contrasted to the string frame)²²

$$G_{E\mu\nu} \stackrel{\text{def}}{=} e^{-\Phi/2} G_{\mu\nu}, \quad (8.76)$$

which is defined by the property that the gravitational term in the action takes the canonical Einstein-Hilbert form

$$\frac{1}{2\kappa^2} \int \sqrt{-G_E} R_E \quad (8.77)$$

without the dilaton factor $e^{-2\Phi}$ in front of the scalar curvature. In other words, the Einstein frame diagonalizes the graviton/dilaton kinetic terms. Using formulae from BOX 1.12 and Eq. (8.46) with $f = \Phi/4$, we get the following dictionary between the string and Einstein frames:

$$\sqrt{-G} e^{-2\Phi} R = \sqrt{-G_E} \left(R_E - \frac{9}{2} \partial^\mu \Phi \partial_\mu \Phi \right) + \text{surface term} \quad (8.78)$$

$$\sqrt{-G} e^{-2\Phi} G^{\mu\nu} \partial_\mu \Phi \partial_\nu \Phi = \sqrt{-G_E} \partial^\mu \Phi \partial_\mu \Phi, \quad (8.79)$$

$$\sqrt{-G} |X^{(p)}|^2 = e^{(5-p)\Phi/2} \sqrt{-G_E} |X^{(p)}|^2, \quad (8.80)$$

where in the RHS indices are raised/contracted with the Einstein frame metric $G_{E\mu\nu}$.

²² Here and below a subscript E means that the quantity is in the Einstein frame.

Consider the terms in S_{IIB} which are quadratic in the derivatives of scalars

$$\begin{aligned}
 & -\frac{1}{4\kappa_{10}^2} \sqrt{-G_E} \left(\partial^\mu \Phi \partial_\mu \Phi + e^{2\phi} \partial^\mu C^{(0)} \partial_\mu C^{(0)} \right) = \\
 & = -\frac{1}{4\kappa_{10}^2} \sqrt{-G_E} \frac{\partial^\mu (C^{(0)} + i e^{-\Phi}) \partial_\mu (C^{(0)} - i e^{-\Phi})}{e^{-2\Phi}} = \\
 & = -\frac{1}{4\kappa_{10}^2} \frac{\partial_\mu \bar{\tau} \partial^\mu \tau}{(\text{Im } \tau)^2}
 \end{aligned} \tag{8.81}$$

where τ is the complex scalar field

$$\tau \stackrel{\text{def}}{=} C^{(0)} + i e^{-\Phi} \in \mathcal{H}, \tag{8.82}$$

Next, let us consider the kinetic terms for the two 2-forms

$$\begin{aligned}
 & -\frac{1}{4\kappa_{10}^2} \sqrt{-G} \left(e^{-2\Phi} |H^{(3)}|^2 + |F^{(3)} - C^{(0)} H^{(3)}|^2 \right) = \\
 & = -\frac{1}{4\kappa_{10}^2} \sqrt{-G_E} e^\phi \left(e^{-2\Phi} |H^{(3)}|^2 + |F^{(3)} - C^{(0)} H^{(3)}|^2 \right) = \\
 & = -\frac{1}{4\kappa_{10}^2} \sqrt{-G_E} [-H^{(3)} F^{(3)}] \frac{1}{e^{-\phi}} \begin{bmatrix} e^{-2\phi} + (C^{(0)})^2 C^{(0)} \\ C^{(0)} & 1 \end{bmatrix} \begin{bmatrix} -H^{(3)} \\ F^{(3)} \end{bmatrix} \\
 & = -\frac{1}{4\kappa_{10}^2} \sqrt{-G_E} \mathcal{M}^{ij} F_i^{(3)} F_j^{(3)}
 \end{aligned} \tag{8.83}$$

where we changed notation and wrote

$$\begin{bmatrix} F_1^{(3)} \\ F_2^{(3)} \end{bmatrix} = \begin{bmatrix} F^{(3)} \\ H^{(3)} \end{bmatrix} \equiv \begin{bmatrix} dC^{(3)} \\ dB^{(3)} \end{bmatrix} \tag{8.84}$$

and set

$$\mathcal{M} = \Omega' S \Omega, \quad \Omega = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \tag{8.85}$$

where $S \equiv \mathcal{M}^{-1}$ is the *symmetric* matrix

$$S = \frac{1}{\text{Im } \tau} \begin{bmatrix} |\tau|^2 & \text{Re } \tau \\ \text{Re } \tau & 1 \end{bmatrix} \in SL(2, \mathbb{R}) \tag{8.86}$$

which defines the $SL(2, \mathbb{R})$ Hodge norm, cf. **BOX 6.5**:

$$\left\| \begin{bmatrix} F_1^{(3)} \\ F_2^{(3)} \end{bmatrix} \right\|_{\text{Hodge}}^2 \equiv \mathcal{M}^{ij} F_i^{(3)} F_j^{(3)} \equiv e^\phi \left(e^{-2\Phi} |H^{(3)}|^2 + |F^{(3)} - C^{(0)} H^{(3)}|^2 \right). \tag{8.87}$$

The Einstein frame bosonic “action” reads

$$S_{\text{IIB}} = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G_E} \left(R_E - \frac{\partial_\mu \bar{\tau} \partial^\mu \tau}{2(\text{Im } \tau)^2} - \frac{1}{2} \mathcal{M}^{ij} F_i^{(3)} F_j^{(3)} - \frac{1}{2} |\tilde{F}^{(5)}|^2 \right) - \frac{1}{8\kappa_{10}^2} \int \epsilon^{ij} C^{(4)} \wedge F_i^{(3)} \wedge F_j^{(3)}. \quad (8.88)$$

This is invariant under the $SL(2, \mathbb{R})$ symmetry

$$\Lambda = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL(2, \mathbb{R}), \quad ad - bc = 1, \quad (8.89)$$

acting as

$$F_i^{(3)'} = \Lambda_i^j F_j^{(3)}, \quad \tilde{F}^{(5)'} = \tilde{F}^{(5)} \quad (8.90)$$

$$\tau' = \frac{a\tau + b}{c\tau + d}, \quad G'_{E\mu\nu} = G_{E\mu\nu}. \quad (8.91)$$

The $SL(2, \mathbb{R})$ invariance of the τ kinetic terms is obvious since the metric in (8.81) is the Poincaré one on $\mathcal{H}$, while the invariance of the $F^{(3)}$ kinetic terms follows from

$$\mathcal{S}(\tau') = \Lambda \mathcal{S}(\tau) \Lambda^t \Leftrightarrow \mathcal{M}(\tau') = (\Lambda^t)^{-1} \mathcal{M}(\tau) \Lambda^{-1}, \quad (8.92)$$

which just expresses the $SL(2, \mathbb{R})$ invariance of the Hodge norm (BOX 6.5). Any given value of τ is invariant under an $SO(2, \mathbb{R})$ subgroup of $SL(2, \mathbb{R})$ so that the scalar fields’ space is the coset

$$SL(2, \mathbb{R})/SO(2, \mathbb{R}) \equiv \mathcal{H} \quad (8.93)$$

that is, the upper half-plane with its Poincaré $SL(2, \mathbb{R})$ -invariant metric. We have the Cartan totally geodesic (isometric) map

$$SL(2, \mathbb{R})/SO(2, \mathbb{R}) \longrightarrow (\text{positive-definite real symmetric } 2 \times 2 \text{ matrices}), \quad (8.94)$$

$$[E] \longmapsto EE^t \equiv \mathcal{S}, \quad (8.95)$$

where $E \in SL(2, \mathbb{R})$ is any representative of the class $[E] \in SL(2, \mathbb{R})/SO(2, \mathbb{R})$ ($\mathcal{S} \equiv EE^t$ is independent of its choice). The Cartan image $\mathcal{S}$ is the matrix in (8.86) which transforms as $\mathcal{S} \rightarrow \Lambda \mathcal{S} \Lambda^t$ for $\Lambda \in SL(2, \mathbb{R})$.

We conclude with a number of fundamental comments:

Remark 8.1 For reason better explained in [1] the fermionic fields are invariant under $SL(2, \mathbb{R})$ and transform only under the compact group $SO(2, \mathbb{R}) \simeq U(1)$.

Remark 8.2 In the physical applications the fluxes of the 3-form field-strengths are quantized in integral units. Indeed by Gauss’ law the $H^{(3)}$ flux measures the total

winding number which is integrally quantized. Because of flux quantization, *only the arithmetic subgroup* $SL(2, \mathbb{Z}) \subset SL(2, \mathbb{R})$ *is an actual symmetry of the model.*

Remark 8.3 The $SL(2, \mathbb{Z})$ symmetry rotates one 2-form field into the other. From the viewpoint of perturbative string theory there seems to be an asymmetry between the two potentials: the NS-NS 2-form gauge-field has a source, the string, whereas we have not yet encountered any extended object which sources the R-R gauge-field. One may suspect that $SL(2, \mathbb{Z})$ is merely an accidental symmetry of the low-energy theory, explicitly broken by massive string modes. But it is not so: going to the non-perturbative completion of the theory, we shall see in Chap. 13 that the arithmetic subgroup $SL(2, \mathbb{Z})$ is an *exact symmetry* of the full IIB string, in fact a *gauge symmetry*.

Remark 8.4 Type IIB supersymmetry, together with flux quantization, fixes the low-energy effective Lagrangian up to the gauging by a subgroup $\Gamma \subset SL(2, \mathbb{Z})$, so the actual scalars' moduli space is the double coset

$$\Gamma \backslash SL(2, \mathbb{R}) / SO(2) \quad (8.96)$$

which has finite volume iff Γ has *finite index* in $SL(2, \mathbb{Z})$ i.e. iff Γ is an arithmetic group, as required by the swampland conjectures. We shall see in Chap. 13 that this property is maximally satisfied in actual Type IIB where $\Gamma \equiv SL(2, \mathbb{Z})$.

8.6 Type I Superstring: Low-Energy Effective Theory

We get Type I SUGRA [48, 49] from Type IIB in 2 steps inspired by string theory:

- set to zero the IIB fields $C^{(0)}$, $B^{(2)}$ and $C^{(4)}$ which are removed by the Ω projection;
- add the gauge fields with the correct dilaton dependence for open-string fields.

In addition we need to modify the $F^{(3)}$ field strength for reasons to be explained later. The resulting bosonic part of action is

$$S_I = S_c + S_o \quad (8.97)$$

$$S_c = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G} \left[e^{-2\Phi} \left(R + 4 \partial^\mu \Phi \partial_\mu \Phi \right) - \frac{1}{2} |\tilde{F}^{(3)}|^2 \right], \quad (8.98)$$

$$S_o = -\frac{1}{2g_{10}^2} \int d^{10}x \sqrt{-G} e^{-\Phi} \text{Tr}_V (|F^{(2)}|^2). \quad (8.99)$$

where we distinguish couplings arising from the closed and open sectors.

The open string $SO(32)$ gauge potential and field strength are written as matrix-valued forms $A^{(1)}$ and $F^{(2)}$, which transform in the vector representation as indicated by the subscript on the trace. $\tilde{F}^{(3)}$ is the *improved* 3-form field strength

BOX 8.4 - Eq. (8.99)

The 10d $\mathcal{N} = 1$ supergravity-Yang Mills action may be read in the original paper [48]. In the *Einstein frame* the bosonic part of the Lagrangian reads ($e \equiv \sqrt{-G_E}$)

$$L = \frac{1}{2}eR - \frac{1}{8}e\phi^{-3/2}|F^{(3)'}|^2 - \frac{1}{2}e\phi^{-3/4}|F^{(2)}|^2 + \text{fermions}$$

Comparing with (8.80), we see that $\phi^{-3/2} \sim e^\Phi$, so that in the *Einstein frame* the coupling in front of the Yang Mills term $|F^{(2)}|^2$ is $e^{\Phi/2}$. Using again (8.80)

$$\sqrt{-G_E} e^{\Phi/2} |F^{(2)}|^2 = \sqrt{-G} e^{-3\Phi/2} e^{\Phi/2} |F^{(2)}|^2 \equiv \sqrt{-G} e^{-\Phi} |F^{(2)}|^2.$$

$$\tilde{F}^{(3)} = dC^{(2)} - \frac{\kappa_{10}^2}{g_{10}^2} \omega^{(3)}, \quad (8.100)$$

with $\omega^{(3)}$ the gauge *Chern-Simons 3-form*

$$\omega^{(3)} \stackrel{\text{def}}{=} \text{Tr}_V \left(A^{(1)} \wedge dA^{(1)} + \frac{2}{3} A^{(1)} \wedge A^{(1)} \wedge A^{(1)} \right). \quad (8.101)$$

Again, the improvement of the field strength reflects a modification of the gauge transformation. Under an ordinary Yang-Mills gauge transformation

$$\delta A^{(1)} = d\lambda - i[A^{(1)}, \lambda], \quad (8.102)$$

the Chern-Simons form transforms as

$$\delta \omega^{(3)} = d \text{Tr}_V (\lambda F^{(2)}), \quad (8.103)$$

thus it must be that $C^{(2)}$ transforms under the Yang-Mills gauge variation as

$$\delta C^{(2)} = \frac{\kappa_{10}^2}{g_{10}^2} \text{Tr}_V (\lambda F^{(2)}). \quad (8.104)$$

The 2-form gauge transformation $\delta C^{(2)} = d\xi^{(1)}$ remains unaffected. The action S_I (including the CS improvement of the 2-form field-strength) is uniquely determined by supersymmetry and the YM gauge group $\mathcal{G} = SO(32)$.

Exercise 8.3 Show that the above action does not contain any free dimensionless parameters, i.e. that all dimensionless parameters can be set to 1 by a shift of a field.

8.7 Heterotic String

The heterotic string has the same supersymmetry as Type I. Since the low-energy Lagrangian is fully fixed by SUSY and the YM group $\mathcal{G}$, it should be the same one. However the heterotic string does not have an open sector nor a R-R one: the dependence on the dilaton should be an overall $e^{-2\Phi}$ for all terms since the full tree-level effective Lagrangian is computed by sphere amplitudes. Thus the bosonic terms in the heterotic low-energy action are

$$S_{\text{het}} = \frac{1}{2\kappa_{10}^2} \int d^{10}x \sqrt{-G} e^{-2\Phi} \left[R + 4 \partial^\mu \Phi \partial_\mu \Phi - \frac{1}{2} |\tilde{H}^{(3)}|^2 + \frac{\kappa_{10}^2}{g_{10}^2} \text{Tr}_V(|F^{(2)}|^2) \right]. \quad (8.105)$$

Here

$$\tilde{H}^{(3)} = dB^{(2)} - \frac{\kappa_{10}^2}{g_{10}^2} \omega^{(3)}, \quad \delta B^{(2)} = \frac{\kappa_{10}^2}{g_{10}^2} \text{Tr}_V(\lambda F^{(2)}), \quad (8.106)$$

are the same expressions as in Type I but we have changed the names of the fields to reflect the fact that now they come from the NS sector.

Because of the uniqueness of the SUSY Lagrangian, the Type I and heterotic actions may differ only by a field redefinition. It is easy to check that the relation is (quantities with index I refer to Type I, those with index h to the heterotic string)

$$G_{I\mu\nu} = e^{-\phi_h} G_{h\mu\nu} \quad \phi_I = -\phi_h \quad (8.107)$$

$$\tilde{F}_I^{(3)} = \tilde{F}_h^{(3)}, \quad A_I^{(1)} = A_h^{(1)}. \quad (8.108)$$

see **BOX 8.5**. We shall fathom the dynamical meaning of this redefinition in Chap. 13.

BOX 8.5 - Eqs. (8.107) and (8.108)

Using **BOX 1.12** and Eq. (8.46) we have

$$\sqrt{-G_I} R_I = e^{-4\Phi_h} \sqrt{-G_h} R_h + \dots \quad \sqrt{-G_I} |F_I^{(p)}|^2 = e^{(p-5)\Phi_h} \sqrt{-G_h} |F_h^{(p)}|^2$$

so that $\sqrt{-G_I} e^{-2\Phi_I} R_I = e^{-2\Phi_h} \sqrt{-G_h} R_h + \dots$ and

$$\sqrt{-G_I} |\tilde{F}_I^{(3)}|^2 = e^{-2\Phi_h} \sqrt{-G_h} |\tilde{F}_h^{(3)}|^2, \quad \sqrt{-G_I} e^{-\Phi_I} |F_I^{(2)}|^2 = e^{-2\Phi_h} \sqrt{-G_h} |F_h^{(2)}|^2$$

Note 8.6 It is convenient to have a uniform normalization of invariants for all gauge groups. The symbol $\text{Tr}_V(t^a t^b)$ stands for $\frac{1}{h^\vee} \text{Tr}_{\text{ad}}(t^a t^b)$ where $h^\vee$ is the dual Coxeter number. For $\mathcal{G} = SO(2n)$ this reduces to the trace in the vector representation.

8.8 BPS Solutions

For reviews of solutions of the SUGRA e.o.m. see e.g. [50–53].²³

General Properties of Solutions

There are two notions of ‘solution’ of the field equations. The *solitonic* solutions are smooth field configurations which obey the e.o.m. everywhere, while the solutions in presence of a *source* satisfy the field equations only outside some submanifold S where the e.o.m. have a delta-function singularity. The second kind of solutions represent the fields produced by a source with support on S : they are actual physical configurations *provided* there exists in the theory a dynamical object which may play the role of the source. It is one of the consistency requirement of Quantum Gravity [55] that such source objects *do* exist for all (reasonable) solutions in the second sense. We shall check this prediction for the 10d superstrings in Chap. 12.

(Black) Charged Branes We are particularly interested in solutions (in either sense) of D -dimensional SUGRA which describe *charged p -branes* i.e. static objects extended in p spatial directions. More precisely, we say that a solution of the field equations is a *p -brane* iff

- (i) it is asymptotic at infinity to flat space;
- (ii) it is invariant under a symmetry group

$$I \supseteq I_p \equiv \mathbb{R} \times \text{Iso}(\mathbb{R}^p) \times SO(D - p - 1), \quad (8.109)$$

with generic isotropy subgroup $J_p \simeq SO(p) \times SO(D - p - 2) \subset I_p$.

Concretely: $\mathbb{R}$ acts by time translations, $\text{Iso}(\mathbb{R}^p)$ is the p -dimensional Euclidean group acting on the spatial coordinates x^a ($a = 1, \dots, p$) parallel to the brane, and $SO(D - p - 1)$ rotates the coordinates y^i normal to the brane, whose world-story is given by the submanifold $y^i = 0$.

A brane is *charged* iff there is a flux of a $(D - p - 2)$ -form field-strength $\mathfrak{F}^{(D-p-2)}$ through the sphere at infinity in the transverse space,

$$S_\infty^{D-p-2} \equiv \{y^i y^i = R^2 \rightarrow \infty\}. \quad (8.110)$$

One distinguishes between electric and magnetic brane charges depending on the specific form of the field strength $\mathfrak{F}^{(D-p-2)}$ (see below).

Restricted to a slice $x^i = \text{const.}$ a brane solution looks like a $(D - p)$ -dimensional charged black hole (BH)—the prototype being the 4d Reissner-Nordström charged black hole [54, 56, 57]: for this reason these objects are called *black branes* [58]. A generalization of Birkhoff’s theorem from General Relativity [54, 56] guarantees a unique solution for given mass M and charge Q per unit volume along the brane. Just as in the Reissner-Nordström BH, the two data M , Q determine the radial positions r_+ , r_- of the horizon and the singularity: for $M/|Q|$ greater than a certain critical

²³ For exact solutions to Einstein’s equations see e.g. [54].

value $(M/|Q|)_c$, we have $r_+ > r_-$ and the solution is a nice black p -brane whose singularity at r_- is hidden behind a horizon at r_+ . When $M/|Q| < (M/|Q|)_c$ there is a *naked singularity*. Penrose's *cosmic censorship hypothesis* [57, 59, 60] states that this cannot happen in a sound physical configuration. The solution which saturates the inequality, $M/|Q| = (M/|Q|)_c$, is called *extremal* and has special properties. In most supersymmetric theories the extremal solutions are BPS configurations which saturate a BPS bound of the form

$$M \geq \left(\frac{M}{|Q|} \right)_c |Q|. \quad (8.111)$$

Thus backgrounds with naked singularities are excluded in SUSY models by the BPS bound. This is one of the many ways in which supersymmetric theories are 'more regular' than their non-SUSY cousins.

In this textbook we are mainly interested in BPS (hence extremal) brane solutions because of their special non-renormalization properties (cf. Sect. 8.2) which allow to study their dynamics well beyond the weak-coupling regime. Consequently we shall work out the extremal BPS brane solutions in some detail, and discuss their non-BPS generalization only briefly in the last paragraph of this chapter. The explicit formulae presented there will make our general remarks about horizons, singularities, and BPS bounds more concrete.

Note 8.7 There are other kinds of BPS solutions besides extremal branes: SUSY compactifications, pp -waves, intersecting branes, etc. Some of these solutions will be discussed in later chapters of this book. See e.g. Sect. 14.6.

Warped Products & Isometries If M and N are two (pseudo-)Riemannian manifolds, with respective metric ds_M^2 and ds_N^2 , and $f: M \rightarrow \mathbb{R}$ is a smooth function, their *warped product* $M \times_f N$ is the manifold $M \times N$ with metric

$$ds^2 = ds_M^2 + f^2 ds_N^2. \quad (8.112)$$

The function f^2 is called the *warp factor*. See **BOX 8.6** for some useful formulae and [12, 61] for more details.

Proposition 8.1 [61] *Let X be a pseudo-Riemannian manifold, and $H \subset \text{Iso}(X)$ a group of isometries with $\dim H = d(d+1)/2$ whose orbits have dimension d . Then $X = M \times_f Y$ with Y a d -dimensional maximal symmetric space with $\text{Iso}(Y) = H$.*

The orbits have dimension d when the generic isotropy subgroup $K \subset H$ has dimension $d(d-1)/2$. E.g. H may be the Poincaré group $\text{Iso}(\mathbb{R}^{d-1,1})$, the Euclidean group $\text{Iso}(\mathbb{R}^d)$, or the anti-de-Sitter group $SO(d-1, 2)$, with respective isotropy subgroups $SO(d-1, 1)$, $SO(d)$, or $SO(d-1, 1)$. The **Proposition** applies to the black p -branes whose isometry group contains $\text{Iso}(\mathbb{R}^p)$ with

$$K \equiv J_p \cap \text{Iso}(\mathbb{R}^p) \simeq SO(p). \quad (8.113)$$

BOX 8.6 - Curvatures in warped products

We consider a warped product $Z \equiv M \times_f N$ of Riemannian manifolds, where M has metric g_{ij} , N metric γ_{ab} and $\dim M = m$, $\dim N = n$. We write x^i, y^a for the coordinates, and $R_{\alpha\beta\gamma\delta}$, R_{ijkl} and P_{abcd} for the Riemann tensor of Z , M , and N respectively. Then [12, 61]

$$\begin{aligned} R_{ijkl} &= R_{ijkl}, & R_{aibj} &= \gamma_{ab} f D_i \partial_j f, \\ R_{abcd} &= f^2 P_{abcd} - f^2 (\gamma_{ad} \gamma_{bc} - \gamma_{ac} \gamma_{bd}) g^{ij} \partial_i f \partial_j f \end{aligned}$$

where D_b is the covariant derivative in M . The Ricci tensor is

$$R_{ij} = R_{ij} - \frac{n}{f} D_i \partial_j f \quad R_{ab} = P_{ab} - (n-1)(g^{ij} \partial_i f \partial_j f) \gamma_{ab} - f (D^i \partial_i f) \gamma_{ab}$$

Writing $f(y) = e^{\sigma(y)}$ the scalar curvature becomes

$$R = R + e^{-2\sigma} P - 2n D^i \partial_i \sigma - n(n+1) \partial^i \sigma \partial_i \sigma$$

Application For the metric $d\tilde{s}^2 = dy^m dy^m + e^{2\sigma(y)} \eta_{\mu\nu} dx^\mu dx^\nu$ on $Z \equiv \mathbb{R}^m \times_{e^\sigma} \mathbb{R}^{n-1,1}$

$$R = -2n \partial_i \partial_i \sigma - n(n+1) \partial_i \sigma \partial_i \sigma \quad R_{ab} = -e^{2\sigma} \gamma_{ab} (\partial_i \partial_i \sigma - n \partial_i \sigma \partial_i \sigma)$$

For a function $B(y^i)$ on Z which depends only on y^i one has

$$D^i \partial_i B = e^{-n\sigma} \partial_i (e^{n\sigma} \partial_i B) = \partial_i \partial_i B + n \partial_i \sigma \partial_i B$$

We make a conformal redefinition of the metric on Z ,

$$ds^2 = e^{2B(y)} d\tilde{s}^2 = e^{2B(y)} dy^m dy^m + e^{2(\sigma(y)+B(y))} \eta_{\mu\nu} dx^\mu dx^\nu$$

Using the formulae for conformal transformations in **BOX 1.12** we get for the scalar curvature

$$\begin{aligned} R_{ds^2} &= e^{-2B} \left(-2n \partial_i \partial_i \sigma - n(n+1) \partial_i \sigma \partial_i \sigma - \right. \\ &\quad \left. - 2(m+n-1) \partial_i \partial_i B - 2n(m+n-1) \partial_i \sigma \partial_i B - (m+n-2)(m+n-1) \partial_i B \partial_i B \right) \end{aligned}$$

BPS Branes

For *extremal* $(d-1)$ -branes in a D -dimensional spacetime, the isometry and generic isotropy groups enhance from (8.109) to

$$I = \text{Iso}(\mathbb{R}^{d-1,1}) \times SO(D-d) \quad J \simeq SO(d-1,1) \times SO(D-d-1), \quad (8.114)$$

i.e. along the brane we have the full d -dimensional Poincaré symmetry. By **Proposition 8.1**, the most general metric with these symmetries is

$$ds^2 = e^{A(r)} \eta_{\mu\nu} dx^\mu dx^\nu + e^{B(r)} dy^m dy^m, \quad \begin{cases} \mu = 0, 1, \dots, d-1, \\ m = d, d+1, \dots, D-1 \end{cases} \quad (8.115)$$

Table 8.2 Dilaton exponent a_k in the k -form kinetic term for $D \geq 10$ SUGRAs

Model	Sector	k	a_k	Model	Sector	k	a_k
11d SUGRA		3	0	IIB SUGRA	RR	0	-2
IIA SUGRA	NS	2	1	IIB SUGRA	NS	2	1
IIA SUGRA	RR	1	-3/2	IIB SUGRA	RR	2	-1
IIA SUGRA	RR	3	-1/2	IIB SUGRA	RR	4	0

where x^μ are coordinates along the brane, y^m are coordinates normal to the brane, $r \equiv \sqrt{y^m y^m}$, and the functions $A(r)$, $B(r)$ vanish at infinity. The other fields should also respect the symmetry I and go to their vacuum value at infinity.

We focus on the half-BPS branes of 11d SUGRA and 10d IIA/IIB SUGRAs²⁴ which preserve 16 supersymmetries. We consider BPS configurations where the only non-trivial fields are the metric (8.115), the dilaton $\Phi(r)$, and *one* out of the several k -form gauge fields.²⁵ This is legitimate since setting to zero some gauge forms is a consistent truncation of SUGRA (cf. Footnote 8). In the Einstein frame the truncated SUGRA action becomes

$$\frac{1}{2\kappa^2} \int \sqrt{-g} \left(R - \frac{1}{2} \partial^\mu \Phi \partial_\mu \Phi - \frac{1}{2} e^{-a_k \Phi} |F^{(k+1)}|^2 \right) d^D x \quad (8.116)$$

where for 11d SUGRA $a_3 \equiv 0$ since there is no scalar. The numerical constants a_k are fixed by SUSY: they can be read from the effective Lagrangians in Sects. 8.4 and 8.5, see Table 8.2. From the table we note the relation

$$a_k^2 = 4 - \frac{2k(D-k-2)}{D-2} \equiv 4 - 2k + \frac{2k^2}{D-2}, \quad (8.117)$$

a consequence of SUSY [52, 53, 62]. For Type II RR form fields this gives

$$a_k = \frac{k-4}{2}, \quad a_{8-k} = -a_k. \quad (8.118)$$

The k -form gauge fields satisfy a generalized Gauss' law: the *electric* e and *magnetic* m charges carried by the $(d-1)$ -brane are²⁶

²⁴ For heterotic/Type I BPS branes see Sect. 13.4.1.

²⁵ In this section the IIB RR field $C^{(0)}$ is seen as a 0-form gauge field rather than a scalar.

²⁶ The charges are canonically normalized so that e is the Noether charge while m is the topological charge with Dirac's normalization. The RR vertices (i.e. the Kähler-Dirac BRST condition) treat the field strengths $F^{(k)}$ and their dual $*F^{(10-k)}$ symmetrically, so in string theory it may be more

$$e = \frac{1}{2\kappa^2} \int_{S^{D-d-1}} * \left(e^{-a_d \phi} F^{(d+1)} \right), \quad m = \int_{S^{D-d-1}} F^{(D-d-1)} \quad (8.119)$$

where, as before, S_∞^{D-d-1} is the sphere at ∞ in transverse space. For a 0-brane (i.e. a particle) in $D = 4$ Eqs. (8.119) are the usual definitions of electric and magnetic charge. In the general case e and m are charges *per unit volume* for the brane extended in the spatial directions $x^1, \dots, x^{d-1}$. In 4d electrodynamics there is an apparent²⁷ asymmetry between the electric and magnetic charges: the electric charge is a conserved *Noether charge* and $d * (e^{-a_1 \phi} F^{(2)})$ is not zero everywhere, but only away from the 0-brane ($\equiv$ charged particle) source i.e. $d * (e^{-a_1 \phi} F^{(2)})$ is proportional to a delta-function with support on the 0-brane. On the contrary the magnetic charge is a *topological charge* proportional to the first Chern class of the $U(1)$ gauge bundle, and the Bianchi identity $dF^{(2)} = 0$ holds everywhere without delta-function singularities. This (apparent) asymmetry applies to all D and d : e is a Noether charge and m is a topological invariant. For a magnetically charged brane the equations of motion hold everywhere, and the solution is typically regular, while for the electrically charged ones the e.o.m. hold up to delta-function source-terms on the support of the brane, and the field configurations are usually singular there. So the magnetic brane solutions are *bona fide* solitons of the effective theory and are often called *solitonic branes*, whereas the electric one are called *elementary branes* since they have the interpretation of the supergravity fields sourced by an ‘elementary’ brane. We stress again that the difference is not intrinsic, but results from a specific choice of fields to describe the propagating d.o.f.: a more democratic formulation, where no a priori choice is made, exists [63, 64] but it is not ‘economic’. Consequently we have two different $\text{ISO}(\mathbb{R}^{d-1,1}) \times SO(D-d)$ symmetric ansatz for the gauge form-fields from this elementary vs. solitonic dichotomy.²⁸ The electric ansatz is given by a global gauge field $A^{(d)}$, while the magnetic one, being topologically non-trivial, cannot be expressed by a global gauge field, and is written in terms of a global field-strength $F^{(\tilde{d}+1)}$ ($\tilde{d} \equiv D-d-2$).

It is convenient to consider a slightly more general ansatz²⁹ where the background has d -dimensional Poincaré symmetry but it is not necessarily rotational invariant in the transverse directions, that is, in Eq. (8.115) we allow A, B, Φ to be generic

natural to normalize the electric and magnetic charges with the same overall numerical coefficient. Various normalization conventions are used in the literature.

²⁷ The asymmetry is only *apparent* because it depends on which d.o.f. we use to describe the physics: electric fields or the dual magnetic ones. Going to the dual description electric and magnetic charges invert their roles. Recall from the Kähler-Dirac formulation of the R-R gauge fields in Chap. 3 that the superstring treats the gauge fields and their duals in a democratic way.

²⁸ For the moment we consider branes carrying only one type of charge.

²⁹ The original rotational-invariant ansatz, being the most general configuration invariant under a bosonic symmetry, is automatically a consistent truncation of SUGRA and we insert that ansatz directly into the action. That our slightly generalized ansatz is still a consistent truncation may be less obvious; however the property may be checked using the Ricci curvatures for a metric conformally equivalent to the warped product $\mathbb{R}^m \times_{e^A} \mathbb{R}^{n-1,1}$ see **BOX 8.6**.

functions of the transverse coordinates y^m (going to constants at infinity). Then the ansatz for the gauge forms are³⁰ (we set $\tilde{d} = D - d - 2$)

$$\begin{aligned} \text{electric} \quad & F_{\mu_1 \dots \mu_d n} = \epsilon_{\mu_1 \dots \mu_d} \partial_n e^{C(y^m)} \\ \text{magnetic} \quad & \frac{1}{2\kappa^2} e^{-a_k \phi} F_{m_1 \dots m_{\tilde{d}+1} n} = \epsilon_{m_1 \dots m_{\tilde{d}+1} n} \partial_n e^{C(y^m)} \end{aligned} \quad (8.120)$$

Note that the electric ansatz satisfies automatically the Bianchi identity $dF^{(d+1)} = 0$, whereas the magnetic ansatz satisfies automatically the e.o.m. $d(*e^{-a_k \phi} F^{(\tilde{d}+1)}) = 0$. Generalized electro-magnetic duality [1] interchanges the two ansatze.

Dirac Quantization of Electric/Magnetic Charges

The brane electric/magnetic charges satisfy a generalized Dirac quantization condition. We recall the original Dirac argument for point charges [65, 66]: consider a magnetic charge m at the origin of a $\mathbb{R}^3$ space. The flux at infinity of the gauge field-strength is the magnetic charge $m \equiv \int_{S_\infty^2} F^{(2)}$. We cannot write $F^{(2)} = dA^{(1)}$ for a globally defined potential $A^{(1)}$ because the integral over a closed surface is non-zero: we can write $F^{(2)} = dA^{(1)}$ everywhere except along a Dirac string ending at the monopole. Now consider an electric charge e moving in this field along a closed path ℓ . Its coupling to the field produces a holonomy phase

$$\exp\left(ie \oint_\ell A^{(1)}\right) = \exp\left(ie \int_D F^{(2)}\right) \quad (8.121)$$

where $\ell = \partial D$. Contract the loop ℓ to a small circle around the Dirac string; the phase (8.121) becomes $\exp(iem)$. The Dirac string must be invisible (since it is a mere gauge artifact) so this phase should be 1. We get the Dirac quantization condition

$$e m = 2\pi n, \quad n \in \mathbb{Z}. \quad (8.122)$$

Consider now two branes of 10d Type II SUGRA charged, respectively, electrically and magnetically for the *same* d -form gauge field: we have an electric $(d-1)$ -brane and a magnetic $(7-d)$ -brane. In the $\mathbb{R}^9$ space the $(7-d)$ -dimensional brane is surrounded (i.e. linked) by a $(d+1)$ -sphere, and its magnetic charge (per unit volume) is

$$m_{7-d} = \int_{S^{d+1}} F^{(d+1)}. \quad (8.123)$$

We take the $(d-1)$ -brane to be extended in the directions $4 \leq \mu \leq d+2$ and the $(7-p)$ -brane in the directions $p+4 \leq \mu \leq 9$. Our two brane system effectively reduces to the 4d situation for charged point-particle in the transverse $\mathbb{R}^3$ of coordinates y^1, y^2, y^3 . Applying the Dirac argument to this set-up, we get that the charges (per unit volume) satisfy

³⁰ Here $\epsilon_{\mu_1 \dots \mu_d}$ is the totally antisymmetric tensor with $\epsilon_{0 \dots d-1} = 1$ and the same for $\epsilon_{m_1 \dots m_{\tilde{d}+1} n}$.

$$em = 2\pi n, \quad n \in \mathbb{Z}. \quad (8.124)$$

Note 8.8 The electric charge e is quantized in integer multiples of some fundamental charge that we may set to 1 as a choice of normalization. With this normalization (12.42) says that the class $[F^{(2)}/2\pi] \in H^2(M, \mathbb{R})$ is *integral*. Indeed $[F^{(2)}/2\pi]$ represents the *first Chern class* $c_1(L) \in H^2(M, \mathbb{Z})$ of the $U(1)$ bundle L where charged particle wave-function takes value [67–71].

BPS Electric Branes Solutions We consider first electrically charged branes. We exploit the fact that our ansatz (8.115), (8.120) is a consistent truncation of SUGRA. Then, using the formulae in BOX 8.6 with $\sigma \equiv A - B$, we get ($\partial_i \equiv \partial_{y^i}$)

$$\begin{aligned} S|_{\text{ansatz}} = \frac{1}{2\kappa^2} \int d^D x \Big\{ & e^{d\sigma + (D-2)B} \Big(d(d-1)\partial_i \sigma \partial_i \sigma + 2d(D-2)\partial_i \sigma \partial_i B + \\ & + (D-1)(D-2)\partial_i B \partial_i B - \frac{1}{2}\partial_i \Phi \partial_i \Phi \Big) + \\ & + \frac{1}{2} e^{-d\sigma + (D-2)B - a_k \Phi + 2C} \partial_i C \partial_i C \Big\} \end{aligned} \quad (8.125)$$

Exercise 8.4 Check that, when the relation (8.117) holds (for $k = d$), we can find solutions, asymptotic to flat space at ∞ , which satisfy the relations³¹ [51–53, 62]

$$\sigma = \frac{1}{2}(C - C_0), \quad B = -\frac{d}{2(D-2)}(C - C_0), \quad a_k(\Phi - \Phi_0) = \frac{a_k^2}{2}(C - C_0) \quad (8.126)$$

where $C_0 = a_k \Phi_0/2$ and Φ_0 are the values at infinity.

Exercise 8.5 Show that, under the same conditions, *all* $SO(D-d)$ -invariant solutions which are asymptotic to Minkowski space satisfy (8.126). This is an instance of the generalized Birkhoff's theorem alluded before.

Equation (8.126) has an important physical interpretation: these are precisely the conditions for the configuration to be $\frac{1}{2}$ -BPS, i.e. $\delta^{(a)}\psi_\mu = \delta^{(a)}\chi = 0$ for 16 supercharges. In BOX 8.7 this claim is illustrated in the case of BPS 2-branes in 11d SUGRA; KK reduction then extends the claim to 10d Type IIA, and the Type IIB case follows from the relation between the two 10d theories.

It follows from Eq. (8.126) that the BPS configurations of 11d and 10d IIA/IIB SUGRA with d -dimensional Poincaré symmetry which go to a Minkowski vacuum at ∞ depend on a single function $C(y^m)$ of the transverse coordinates. We remain with one last e.o.m. for $C(y^m)$. Varying the action (8.125), and imposing the conditions (8.126), we get

$$\left. \frac{\delta S}{\delta C} \right|_{(8.126)} \equiv \partial_i \partial_i C - \partial_i C \partial_i C = 0 \quad \Rightarrow \quad \partial_i \partial_i e^{-C} = 0, \quad (8.127)$$

³¹ Note that the two exponents in (8.125) $d\sigma + (D-2)B$ and $-d\sigma + (D-2)B - a_k \phi + 2C$ vanish when (8.126) holds. This fact greatly simplifies the algebraic manipulations.

BOX 8.7 - $\frac{1}{2}$ -BPS 2-Branes in 11d SUGRA

We consider the 11d SUGRA bosonic background [72] ($\mu = 0, 1, 2$ and $i = 1, \dots, 8$)

$$ds^2 = e^{2A(y)} \eta_{\mu\nu} dx^\mu dx^\nu + e^{2B(y)} \delta_{ij} dy^i dy^j, \quad C_{\mu\nu\rho} = \pm \epsilon_{\mu\nu\rho} e^{C(y)}$$

The gravitino SUSY variation in a bosonic background is

$$\delta\psi_M|_{\text{bos}} = \left(\partial_M + \frac{1}{4} \omega_M^{AB} \Gamma_{AB} - \frac{1}{288} (\Gamma^{PQR} S_M + 8 \Gamma^{PQR} \delta^S_M) F_{PQRS} \right) \varepsilon$$

where $F_{MNPQ} \equiv 4\partial_{[M} C_{NPQ]}$ and Γ_A are 11d Dirac matrices that we choose as

$$\Gamma_\alpha = \gamma_\alpha \otimes \Gamma_9 \quad (\alpha = 0, 1, 2), \quad \Gamma_{a+2} = \mathbf{1} \otimes \Sigma_a \quad (a = 1, \dots, 8),$$

with $\Gamma_9 = \Sigma_1 \Sigma_2 \dots \Sigma_8$ the chirality operator in 8d, and $\varepsilon = \epsilon \otimes \eta(r)$ with ϵ (resp. η) a constant $\text{Spin}(1, 2)$ spinor (resp. a y^i -dependent $\text{Spin}(8)$ spinor). Using the vielbeins $e^\mu = e^A dx^\mu$, $e^i = e^B dy^i$ (where boldface indices are ‘flat’) we get the spin-connection [62]

$$\omega^{\mu\nu} = 0, \quad \omega^{\mu i} = e^{-B} \partial_i e^\mu, \quad \omega^{ij} = e^{-B} (\partial_j B) e^i - e^{-B} (\partial_i B) e^j$$

and the BPS conditions become

$$\begin{aligned} (\partial_\mu - \frac{1}{2} \gamma_\mu e^{-A} \Sigma^i \partial_i e^A \Gamma_9 \mp \frac{1}{6} \gamma_\mu e^{-3A} \Sigma^i \partial_i e^C) \epsilon \otimes \eta &= 0 \\ (\partial_i + \frac{1}{4} e^{-B} (\Sigma_i \Sigma^j - \Sigma^j \Sigma_i) \partial_j e^B \mp \frac{1}{24} e^{-3A} (\Sigma_i \Sigma^j - \Sigma^j \Sigma_i) \partial_i e^C \Gamma_9 \mp \frac{1}{6} e^{-3A} \partial_i e^C \Gamma_9) \epsilon \otimes \eta &= 0 \end{aligned}$$

Choosing η to be a chiral 8d spinor, $\Gamma_9 \eta = \mp \eta$, the first equation is satisfied iff $C = 3A$; the terms proportional to $(\Sigma_i \Sigma^j - \Sigma^j \Sigma_i)$ cancels iff $C = -6B + \text{const}$. We remain with the equation

$$(\partial_i + \frac{1}{6} e^{-C} \partial_i e^C) \eta = 0$$

whose solution is $\eta(y) = e^{-C(y)/6} \eta_0$ with η_0 a constant spinor. Hence [72]

$$\sigma \equiv A - B = \frac{1}{2}(C - C_0), \quad B = -\frac{1}{6}(C - C_0) = -\frac{3}{2(11-2)}(C - C_0)$$

that is, $H(y^m) \equiv e^{C_0 - C(y^m)}$ is a harmonic function of the transverse coordinates y^m . If we assume rotational invariance, H becomes (up to an additive constant) the Coulomb potential in $D - d$ dimensions

$$H(r) \equiv e^{-(C-C_0)} = \begin{cases} 1 + \frac{\lambda_d}{r^{D-d-2}} & D - d \neq 2 \\ 1 - \lambda_{D-2} \log y & D - d = 2, \end{cases} \quad (8.128)$$

where λ_d is a constant proportional to the electric charge per unit volume of the brane. In particular the rotational-invariant solution is unique for a given value of the charge. Clearly the RHS of (8.127) is not zero, but rather a delta-function with support at the origin, as expected for a ‘fundamental’ solution.

The magnetic brane is obtained from the electric one by the replacements

$$F^{(d+1)} \leftrightarrow \frac{1}{2\kappa^2} e^{-a_{D-d-2}\Phi} * F^{(D-d-1)}, \quad a_d \leftrightarrow a_{D-d-1} \equiv -a_d. \quad (8.129)$$

The e.o.m. get mapped into the Bianchi identity, so now there is no source term.

Half-BPS Branes Carrying a RR d -Form Charge We specialize our findings to a BPS $(d-1)$ -brane of 10d Type II SUGRA which is electrically charged for the RR d -form gauge field. In the Einstein frame the $\frac{1}{2}$ -BPS solution is

$$ds_E^2 = H^{\frac{d}{8}-1} \eta_{\mu\nu} dx^\mu dx^\nu + H^{\frac{d}{8}} dy^i dy^j \quad (8.130)$$

$$e^{2\Phi} = e^{2\Phi_0} H^{(4-d)/2}, \quad A^{(d)} = \epsilon^{(d)} e^{a_d \Phi_0/2} H^{-1} \quad (8.131)$$

with H as in (8.128) and $\epsilon^{(d)} \equiv dx^0 \wedge \dots \wedge dx^{d-1}$. In the string frame the metric is

$$ds^2 = e^{\Phi/2} ds_E^2 = H^{-1/2} \eta_{\mu\nu} dx^\mu dx^\nu + H^{1/2} \delta_{ij} dy^i dy^j \quad (8.132)$$

For future reference, we add some comments on the solution.

Remark 8.5 The relation of the constant λ_d to the R-R charge per unit volume e can be obtained by comparing with the $(10-d)$ -dimensional Coulomb law. Writing $\text{Vol}(S^{9-d})$ for the volume of the sphere S^{9-d} , we have [53]

$$\lambda_d = \frac{2\kappa^2 e^{C_0} |e|}{(8-d) \text{Vol}(S^{9-d})} \equiv 2\kappa^2 e^{C_0} |e| \pi^{(10-d)/2} \Gamma\left(\frac{8-d}{2}\right). \quad (8.133)$$

Remark 8.6 The above rotational-invariant solution describes an isolated BPS $(d-1)$ -brane at $y^i = 0$. Since the surviving equation of motion (RHS of (8.127)) is linear in $H \equiv e^{C_0-C}$, the superposition of the solutions describing several parallel $(d-1)$ -branes, all extended in the directions $x^1, \dots, x^{d-1}$, but at different positions $y_{(s)}^i$ in transverse space, is still an exact BPS solution described by Eqs. (8.130) and (8.131) with $H(y^i)$ the harmonic function

$$H(y^i) = 1 + \sum_s \frac{\lambda_d}{|\mathbf{y} - \mathbf{y}_s|^{8-d}}. \quad (8.134)$$

Therefore any configuration of $\frac{1}{2}$ -BPS branes, all of the same dimension, parallel and at relative rest, is also $\frac{1}{2}$ -BPS regardless of the separations $|\mathbf{y}_s - \mathbf{y}_t|$ between them. Such a configuration saturates the BPS bound, and since the electric/magnetic charges are additive, its total mass is $\sum_s M_s$ for all transverse separations: this entails that there is *no* net force between parallel equidimensional branes at rest. We shall understand the physical origin of this surprising result in Chap. 12.

Remark 8.7 Consider (say) the BPS 2-brane solution of 11d SUGRA (cf. BOX 8.7). We may compactify 11d SUGRA to 10d Type IIA by identifying periodically the

coordinate x^2 along the brane. In this way we get *simultaneous* compactifications of space-time and brane world-volume. We end up with a 1-brane (i.e. a *string*) solution of 10d Type IIA SUGRA. This ‘elementary’ solution represents the massless fields sourced by a fundamental Type IIA string.

Remark 8.8 We can consider more general solutions where more than one R-R charge is non-zero. The BPS conditions become slightly subtler, and we prefer to discuss them from a higher point of view in Chap. 14.

The Magnetic NS 5-Brane Soliton

We see from Table 8.2 that a NS-NS gauge form-field has a constant a_k of opposite sign with respect to the RR gauge field of the same degree k . This change of sign has dramatic consequences: from (8.126) we see that it flips the sign of the dilaton, i.e. interchanges strong and weak string coupling.

We focus on the $\frac{1}{2}$ -BPS brane which is magnetically charged for the NS-NS 2-form field $B_{\mu\nu}$. It is a 5-brane ($d = 6$). The Einstein frame solution is given by (8.126), (8.128) with $a_6 = -1$ (*opposite sign* with respect to the RR charged 5-brane):

$$ds_E^2 = H^{-1/4} \eta_{\mu\nu} dx^\mu dx^\nu + H^{3/4} \delta_{ij} dy^i dy^j, \quad (8.135)$$

$$e^{2(\Phi - \Phi_0)} = H, \quad H_{ijk} = e^{-\Phi_0/2} \epsilon_{ijkl} \partial_l H^{-1} \quad (8.136)$$

where

$$H = 1 + \frac{\lambda_6}{r^2}, \quad \lambda_6 = 2\kappa^2 \pi^2 e^{C_0} > 0. \quad (8.137)$$

The net effect of the sign flip is to invert the sign of the dilaton, cf. (8.126). Note that $H^{-1} = r^2/(r^2 + \lambda_6)$ is a smooth function in $\mathbb{R}^4$, so the gauge field is topologically non-trivial and everywhere regular. Because of the sign flip, the string frame metric now reads³²

$$ds^2 = e^{\Phi/2} ds_E^2 = \eta_{\mu\nu} dx^\mu dx^\nu + H(r) \delta_{ij} dy^i dy^j, \quad (8.138)$$

i.e. it is the Riemannian (*not* warped!) product $\mathbb{R}^{5,1} \times M_4$ where M_4 has metric

$$ds_{M_4}^2 = \left(1 + \frac{\lambda_6}{r^2}\right) \left(dr^2 + r^2 d\Omega_3^2\right), \quad (8.139)$$

with $d\Omega_3^2$ the round metric on the 3-sphere. The geometry of the 4-manifold M_4 is shown in Fig. 8.1. It has two infinite ends: one is asymptotic to the standard flat metric on $\mathbb{R}^4$; the other one is an infinite cylindrical *throat*. The point $r = 0$ is at infinite distance, and, as one approaches it, the radius of the angular 3-sphere does not shrink to zero but approaches the asymptotic value $\sqrt{\lambda_6}$. M_4 is geodesically complete, so, in the string frame, this BPS solution is *regular everywhere* and hence a *bona fide*

³² We absorb an overall constant in the normalization of the coordinates.

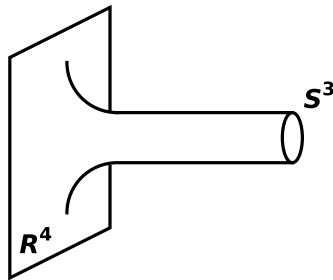


Fig. 8.1 The non-flat factor manifold M_4 in the string-frame geometry of the $\frac{1}{2}$ -BPS 5-brane soliton magnetically charged under the NS-NS gauge field $B_{\mu\nu}$. M_4 is geodesically complete and has two infinite ends: one is asymptotically isometric to $\mathbb{R}^4$ and one to the cylinder $\mathbb{R} \times S^3$ (the ‘throat’)

supergravity soliton: this solution is usually called the *NS5 soliton*. The dilaton $\Phi(r)$ grows linearly with the distance ℓ along the throat

$$\Phi = \frac{1}{2} \log H + \text{const} \approx -\log r + \text{const} \approx \ell \quad \text{as } r \rightarrow 0. \quad (8.140)$$

Therefore string perturbation theory breaks down at some distance down in the throat, and the NS5 soliton is really a non-perturbative object of string theory. We can nevertheless study it in some detail because it preserves 16 supersymmetries.

The geometry is rather different in the Einstein-frame metric ds_E^2 . In the Einstein metric the radial distance is

$$ds_E \sim r^{-3/4} dr, \quad (8.141)$$

so the singularity $r = 0$ is at finite distance and the Einstein metric is not regular. We see another example of the Quantum Gravity fact that different probes detect very different spacetime geometries.

Just as in the case of BPS branes charged under RR fields, when we have a set of parallel NS5 solitons at relative rest, the full system is BPS and the branes do not exert forces between them.

We can look at fluctuations of the fields around the classical NS5 background. The light modes of the fluctuations will represent dynamical d.o.f. living on the NS5 brane. The analysis depends on whether we are in Type IIA or in Type IIB. The solution itself holds equally well for both string models, since only NS-NS fields have a non-trivial background in the NS5 solution, and the (consistent) truncations of the e.o.m. to the NS-NS sector coincide for the two effective theories, but the light modes of R-R and Fermi fields in this background evidently depend on the Type of theory. We shall describe the d.o.f. living on the NS5 brane in Sect. 13.6.

Exercise 8.6 Compute the mass per unit volume ($\equiv$ tension) of the NS5 soliton.

From this **Exercise** the reader will learn that the NS5 tensions is proportional to g^{-2} , where g is the string coupling constant. Hence this BPS object becomes

infinitely massive in the perturbative limit $g \approx 0$, which means that the NS5 soliton is a truly non-perturbative feature of Type II superstring.

Exercise 8.7 Show that the Type IIA NS5 soliton uplifts to a $\frac{1}{2}$ -BPS magnetic 5-brane soliton of 11d SUGRA (inverting the KK reduction).

Relation with Type IIB $SL(2, \mathbb{Z})$ Duality

Having found a BPS solution of SUGRA, we may generate other BPS solutions by acting on the original solution with a group of symmetries of the theory. Of particular interest is the $SL(2, \mathbb{Z})$ duality of 10d Type IIB SUGRA. This symmetry rotates the field strengths $H^{(3)}$ and $F^{(3)}$ of the NS-NS and R-R 2-form gauge fields, preserving the Dirac quantization of the respective fluxes, while acting by modular transformations on the complex scalar τ .

The element $S \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in SL(2, \mathbb{Z})$ transforms a 1-brane (resp. a 5-brane) electrically (resp. magnetically) charged under the R-R 2-form field $A_{\mu\nu}$ into a 1-brane (resp. a 5-brane) electrically (resp. magnetically) charged under the NS-NS 2-form field $B_{\mu\nu}$, while flipping the sign of Φ (setting $C^{(0)} = 0$ for simplicity). The 5-brane solution (8.130)–(8.132) is mapped into the 5-brane soliton (8.135)–(8.138). The $SL(2, \mathbb{Z})$ -orbit of this solution consist of $\frac{1}{2}$ -BPS 5-branes which carry p Dirac units of magnetic NS-NS charge and q Dirac units of R-R magnetic charge with (p, q) a pair of coprime integers.³³ In the same vein we get BPS strings ($\equiv$ 1-branes) with NS-NS and R-R electric charges (p, q) for all coprime integers p, q .

Black Branes

We fulfil our promises, and consider (non-BPS) *black branes*, i.e. objects extended in $p \leq 6$ spatial directions which look as magnetically charged black holes in the transverse directions [58]. We look for a solution with the symmetry in I_p in Eq. (8.109). By a simple generalization of the previous steps one finds [50, 53, 73]

$$ds^2 = -W(r)^{-1}H(r)^{\frac{p+1}{8}} dt^2 + H(r)^{\frac{p+1}{8}} \left(\frac{W(r)}{H(r)} dr^2 + r^2 d\Omega_{8-p}^2 \right) + H(r)^{-\frac{7-p}{8}} dx^i dx^i \quad (8.142)$$

$$e^{2\phi(r)} = H(r)^{a_{7-p}}, \quad F^{(8-p)} = (7-p)(r_+ r_-)^{\frac{7-p}{2}} \omega_{8-p} \quad (8.143)$$

where $d\Omega_{8-p}^2$ and ω_{8-p} are the round metric and volume form on S^{8-p} , while

$$H(r) = 1 + \frac{r_-^{7-p}}{r^{7-p}}, \quad W(r) = 1 + \frac{r_+^{7-p}}{r^{7-p} - (r_+^{7-p} - r_-^{7-p})}. \quad (8.144)$$

³³ The pairs (p, q) of coprime integers, taken modulo PCT, i.e. $(p, q) \sim (-p, -q)$, are in 1-to-1 correspondence with the points in the projective line over the rational $\mathbb{P}^1(\mathbb{Q})$. Clearly $\mathbb{P}^1(\mathbb{Q})$ is the modular group $PSL(2, \mathbb{Z})$ modulo its unipotent subgroup $T^{\mathbb{Z}}, \mathbb{P}(\mathbb{Q}) \simeq PSL(2, \mathbb{Z})/T^{\mathbb{Z}}$.

In these coordinates the two horizons are at $r_1 = 0$ and $r_2 = (r_+^{7-p} - r_-^{7-p})^{1/(7-p)}$ and the singularity as $r_1 = 0$ is ‘clothed’ by the horizon iff $r_+ \geq r_-$. The magnetic charge and mass (per unit volume) of this solutions are [73]

$$q = \frac{\text{Vol}(S^{8-p})}{2\kappa^2} (7-p) (r_+ r_-)^{\frac{7-p}{2}}, \quad (8.145)$$

$$M = \frac{\text{Vol}(S^{8-p})}{2\kappa^2} \left((7-p) r_+^{7-p} + (r_+^{7-p} - r_-^{7-p}) \right) \geq |q| \quad (8.146)$$

with equality iff $r_+ = r_-$, which correspond to the *extremal* black p -brane, which is the $\frac{1}{2}$ -BPS brane studied before.

★ BPS pp -Waves & the Penrose Limit

Generalized pp -waves are pseudo-Riemannian geometries in Lorentzian signature $(d-1, 1)$ whose holonomy group is *non-reductive*.³⁴ When complete and simply-connected, they can be written as products $X \times Y$ where Y is a Riemannian manifold and X a Lorentzian-signature space whose holonomy is indecomposable but not irreducible.³⁵ This implies that a light-like *line* in TX is preserved by parallel transport. A Brinkmann space is an indecomposable Lorentzian manifold which satisfies the stronger condition of having a parallel light-like *vector field* v [78] which is automatically Killing. A non-stationary BPS background in SUGRA has the form $X \times Y$ with X a Brinkmann space and Y of special holonomy (cf. Chap. 11). A distinguished class of Brinkmann spaces are the (indecomposable) *Lorentzian symmetric spaces* [79] whose holonomy group is necessarily solvable; in harmonic coordinates their metric reads

$$ds_X^2 = 2 dx^+ dx^- + A_{ij} x^i x^j (dx^-)^2 + dx^i dx^i \quad i = 2, 3, \dots, \dim X - 1, \quad (8.147)$$

where A_{ij} is a non-degenerate symmetric matrix. We can construct a Brinkmann solution to the e.o.m. of GR from any solution by taking a *Penrose limit* [80]: one zooms on a suitably infinitesimal neighborhood of a segment ℓ of light-like geodesic which does not contain pairs of conjugate points: the tangent vectors to ℓ play the role of the Brinkmann parallel vector field v . This procedure has been extended from the metric to all the fields of SUGRA [81–84].

There is a large family of BPS solutions to 11d and 10d IIA/IIB SUGRAs whose target space M is a Brinkmann space with non-zero gauge-form fluxes: see [85, 86]. The BPS solutions with M a symmetric space preserve generically 16 supersymmetries, but there is a special one which preserves 32—so it is maximally symmetric—As an example we write down the maximal ones, referring the reader to the cited literature for further details. The maximally symmetric pp -wave solutions have metrics [87–89]³⁶

$$\text{11d SUGRA} \quad ds^2 = 2 dx^+ dx^- - \left(\sum_{i=1}^3 x^i x^i + \frac{1}{4} \sum_{i=1}^9 x^i x^i \right) (dx^-)^2 + dx^i dx^i \quad (8.148)$$

$$\text{10d IIB} \quad ds^2 = 2 dx^+ dx^- - (x^i x^i) (dx^-)^2 + dx^i dx^i, \quad (8.149)$$

and gauge-form field-strengths [87–89]

³⁴ This is a situation which happens only in Lorentzian signature: the holonomy group of a positive-signature metric is always reductive being a closed subgroup of the compact group $O(d)$.

³⁵ For the theory of Lorentzian holonomy groups see [74–77]; for a review Sect. 3.7 of [1].

³⁶ $i = 1, \dots, 9$ for 11d and $i = 1, \dots, 8$ for 10d IIB.

$$\text{11d SUGRA} \quad F^{(4)} = 3 dx^- \wedge dx^1 \wedge dx^2 \wedge dx^3, \quad (8.150)$$

$$\text{10d IIB} \quad F^{(5)} = \frac{1}{2} dx^- \wedge (dx^1 \wedge dx^2 \wedge dx^3 \wedge dx^4 + dx^5 \wedge dx^6 \wedge dx^7 \wedge dx^8) \quad (8.151)$$

The solution for 10d IIA SUGRA can be obtained by dimensional reduction of the 11d one. These solutions have 32 supersymmetries [87–89].

Since Brinkmann spaces have a light-like parallel (hence Killing) vector field $\partial/\partial x^+$, strings moving in this background have a nice light-cone quantization. The R-R background is non-trivial, so they are most conveniently described in the Green-Schwarz formulation, which is hard to quantize in a covariant gauge but relatively straightforward in the light-cone one. This leads to a world-sheet description of the superstring in these background which is quite elegant; the details are outside the scope of this book: the readers are referred to [85–91].

References

1. Cecotti, S.: *Supersymmetric Field Theories*. Cambridge University Press, Geometric Structures and Dualities (2015)
2. Freedman, D.Z., van Nieuwenhuizen, P., Ferrara, S.: Progress toward a theory of supergravity. *Phys. Rev. D*. **13**, 3214–3218 (1976)
3. Deser, S., Zumino, B.: Consistent supergravity. *Phys. Lett. B*. **62**, 335 (1976)
4. Freedman, D., van Proeyen, A.: *Supergravity*, CUP 2012
5. Salam, A., Sezgin, E. (eds.): *Supergravities in Diverse Dimensions*, Vol. 1, 2. World-Scientific, North-Holland (1989)
6. Castellani, L., D’Auria, R., Fré, P.: *Supergravity and Superstrings: A Geometric Perspective*. World Scientific (1991)
7. Fré, P.G.: *Advances in geometry and Lie algebras from supergravity*. Springer (2018)
8. Atiyah, M.F., Bott, R., Shapiro, A.: Clifford modules. *Topol.* **3**, 3–38 (1964)
9. Strathdee, J.: Extended poincaré supersymmetry. *Int. J. Mod. Phys. A*. **2**, 273–300 (1987)
10. Haag, R., Łopuszanski, J.T., Sohnius, M.: All possible generators of supersymmetries of the S-matrix. *Nucl. Phys. B*. **88**, 257 (1975)
11. Coleman, S.R., Mandula, J.: All possible symmetries of the S matrix. *Phys. Rev.* **159**, 1251–1256 (1967)
12. Besse, A.: *Einstein Manifolds*. Springer (1987)
13. Joyce, D.D.: *Compact Manifolds with Special Holonomy*. Oxford University Press, Oxford Mathematical Monographs (2000)
14. Joyce, D.D.: *Riemannian holonomy groups and calibrated geometry*, Oxford Graduate Texts in Mathematics vol. 12. Oxford University Press (2007)
15. Zumino, B.: Supersymmetry and Kahler manifolds. *Phys. Lett. B*. **87**, 203 (1979)
16. Cremmer, E., Ferrara, S., Girardello, L., Van Proeyen, A.: Yang-Mills theories with local supersymmetry: Lagrangian, transformation laws and super-Higgs effect. *Nucl. Phys. B*. **212**, 413 (1983)
17. Alvarez-Gaume, L., Freedman, D.Z.: Geometrical structure and ultraviolet finiteness in the supersymmetric sigma model. *Commun. Math. Phys.* **80**, 443 (1981)
18. Bagger, J., Witten, E.: Matter couplings in $N = 2$ supergravity. *Nucl. Phys. B*. **222**, 1 (1983)
19. de Wit, B., Tollsten, A.K., Nicolai, H.: Locally supersymmetric $D = 3$ nonlinear sigma models. *Nucl. Phys. B*. **392**, 3–38 (1993)
20. de Wit, B., Berger, I., Samtleben, H.: Gauged locally supersymmetric $D = 3$ nonlinear sigma models. *Nucl. Phys. B*. **671**, 175–216 (2003)
21. Berger, M.: Sur les groupes d’holonomie des variétés à connexion affine et des variétés riemanniennes. *Bull. Soc. Math. Fr.* **83**, 279–330 (1955)

22. Simons, J.: On transitivity of holonomy systems. *Ann. Math.* **76**, 213–234 (1962)
23. Helgason, S.: *Differential Geometry*. Academic Press, Lie groups and symmetric spaces (1978)
24. Cremmer, E., Julia, B.: The $SO(8)$ supergravity. *Nucl. Phys. B.* **159**, 141–212 (1979)
25. Witten, E.: Constraints on supersymmetry breaking. *Nucl. Phys. B.* **202**, 253–316 (1982)
26. Gibbons, G.W., Hull, C.M., Warner, N.P.: The stability of gauged supergravity. *Nucl. Phys. B.* **218**, 173 (1983)
27. Hull, C.M.: The positivity of gravitational energy and global supersymmetry. *Commun. Math. Phys.* **90**, 545 (1983)
28. Nester, J.A.: A new gravitational energy expression with a simple positivity proof. *Phys. Lett. A.* **83**, 241–242 (1981)
29. Nester, J.A.: Positivity of the Bondi gravitational mass. *Phys. Lett. A.* **85**, 259–260 (1981)
30. Witten, E.: A new proof of the positive energy theorem. *Commun. Math. Phys.* **80**, 381–402 (1981)
31. Weinberg, S., Witten, E.: Limits on massless particles. *Phys. Lett. B.* **96**, 59–62 (1980)
32. Postnikov, M.M.: *Lectures in geometry: Lie groups and Lie algebras*. Editorial URSS (1994)
33. Cremmer, E., Julia, B., Scherk, J.: Supergravity theory in 11 dimensions. *Phys. Lett. B.* **76**, 409–412 (1978)
34. Campbell, L.C., West, P.C.: $N = 2$ $d = 10$ nonchiral supergravity and its spontaneous compactification. *Nucl. Phys. B.* **243**, 112–124 (1984)
35. Giani, F., Pernici, M.: $N = 2$ supergravity in ten dimensions. *Phys. Rev. D.* **30**, 325–333 (1984)
36. Howe, P., West, P.C.: The complete $N = 2$ $d = 10$ supergravity. *Nucl. Phys. B.* **238**, 181–219 (1984)
37. Chapline, G., Manton, N.S.: Unification of Yang-Mills theory and supergravity in ten dimensions. *Phys. Lett. B.* **120**, 124–134 (1983)
38. Huq, M., Namazie, M.A.: Kaluza-Klein supergravity in ten dimensions. *Class. Quantum Grav.* **2**, 293–308 (1985)
39. Duff, M.J., Nilsson, B.E.W., Pope, C.N.: Kaluza-Klein supergravity. *Phys. Rep.* **130**, 1–142 (1986)
40. Romans, L.J.: Massive $N = 2A$ supergravity in ten dimensions. *Phys. Lett. B.* **169**, 374–380 (1986)
41. Coleman, S.R.: More about the massive Schwinger model. *Ann. Phys.* **101**, 239 (1976)
42. Tsimpis, D.: Massive IIA supergravities. *JHEP.* **10**, 057 (2005). [arXiv:hep-th/0508214](https://arxiv.org/abs/hep-th/0508214)
43. Schwarz, J.H.: Covariant field equations of chiral $N = 2$ $D = 10$ supergravity. *Nucl. Phys. B.* **266**, 269–288 (1983)
44. Howe, P., West, P.C.: The complete $N = 2$, $d = 10$ supergravity. *Nucl. Phys.* **B238**, 181–219 (1984)
45. Bergshoeff, E., Hull, C., Ortín, T.: Duality in the type-II superstring effective action. *Nucl. Phys. B.* **451**, 547–575 (1995)
46. Bergshoeff, E., Boonstra, H.J., Ortín, T.: S duality and dyonic p -branes in type II string theory. *Phys. Rev. D.* **53**, 7206–7212 (1996)
47. Belov, D., Moore, G.W.: Holographic action for the self-dual field. [arXiv:hep-th/0605038](https://arxiv.org/abs/hep-th/0605038)
48. Bergshoeff, E., de Roo, M., de Wit, B., van Nieuwenhuizen, P.: Ten-dimensional Maxwell-Einstein supergravity, its currents, and the issue of auxiliary fields. *Nucl. Phys. B.* **195**, 97–136 (1982)
49. Chapline, G., Manton, N.S.: Unification of Yang-Mills theory and supergravity in ten dimensions. *Phys. Lett.* **120B**, 105–109 (1983)
50. Ortín, T.: *Gravity and strings*, Cambridge Monographs on Mathematical Physics. Cambridge University Press (2015)
51. M. Duff, Supermembranes, in *Fields, strings, and duality, TASI 1996* eds. C. Efthimiou & B. Greene pp. 219–220, World-Scientific 1997, [arXiv:hep-th/9611203](https://arxiv.org/abs/hep-th/9611203)
52. Stelle, K.S.: BPS Branes in supergravity. [arXiv:hep-th/9803116](https://arxiv.org/abs/hep-th/9803116)
53. Duff, M.J., Khuri, R.R., Lu, J.X.: String solitons. *Phys. Rep.* **259**, 213–325 (1995). [arXiv:hep-th/9412184](https://arxiv.org/abs/hep-th/9412184)

54. Stephani, H., Kramer, D., Maccallum, M., Hoenselaers, C., Herlt, E.: Exact solutions to Einstein's field equations, Second edition. Cambridge University Press (2003)
55. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](#)
56. D'Inverno, R.: Introducing Einstein's Relativity. Oxford University Press (1992)
57. Wald, R.M.: General Relativity. University of Chicago Press (1984)
58. Horowitz, G.T., Strominger, A.: Black strings and p -branes. Nucl. Phys. **B360**, 197 (1991)
59. Penrose, R.: Singularities and time-asymmetry. In: Hawking, S.W., Israel, W. (eds.) General Relativity. An Einstein centenary survey. Cambridge University Press (1979)
60. Penrose, R.: The question of cosmic censorship. J. Astrophys. Astron. **20**, 233–248 (1999)
61. Prvanović, M.: On warped product manifolds. In: Conference "Filomat '94" (Niš, 1994). Filomat **9** part 2 (1995) 169–185
62. Lu, H., Pope, C.N., Sezgin, E., Stelle, K.S.: Stainless super p -branes. Nucl. Phys. B. **456**, 669–698 (1995). [arXiv:hep-th/9508042](#)
63. Bergshoeff, E., Kallosh, R., Ortin, T., Roest, D., van Proeyen, A.: New formulations of $D = 10$ supersymmetry and D8–O8 domain walls. Class. Quantum. Grav. **18**, 3359–3382 (2001). [arXiv:hep-th/0103233](#)
64. Fukuma, M., Oda, T., Tanaka, H.: Comments on T-Dualities of Ramond-Ramond potentials on tori. Prog. Theor. Phys. **103**, 425–466 (2000). [arXiv:hep-th/9907132](#)
65. Dirac, P.A.M.: Quantised singularities in the electromagnetic field. Proc. Roy. Soc. Lond. A. **133**, 60–72 (1931)
66. Coleman, S.R.: The magnetic monopole fifty years later, Proceedings Les Houches Summer School in Theoretical Physics: Gauge Theories in High Energy Physics. Les Houches **1982**, 461–552 (1981)
67. Chern, S.S.: Complex manifolds without potential theory (with an Appendix in the Geometry of Characteristic Classes). Springer (1979)
68. Griffiths, P., Harris, J.: Principles of Algebraic Geometry. Wiley (1978)
69. Hirzebruch, F.: Topological methods in Algebraic Geometry, Classical in Mathematics. Springer (1995)
70. Husemoller, D.: Fibre Bundles, third edition, Graduate Texts in Mathematics **20**. Springer (1994)
71. Milnor, J.W., Stasheff, J.D.: Characteristic Classes. Princeton University Press (1974)
72. Duff, M.J., Stelle, K.S.: Multi-membrane solutions of $D = 11$ supergravity. Phys. Lett. B. **253**, 113–118 (1991)
73. Lu, J.X.: ADM masses for black strings and p -branes. Phys. Lett. B. **313**, 29–34 (1993)
74. Berard-Bergery, L., Ikenakhen, A.: On the holonomy of Lorentzian manifolds. In: Differential Geometry: Geometry in Mathematical Physics and related topics. American Mathematical Society (1993)
75. Galaev, A.: Isometry groups of Lobachewskian spaces, similarity transformations of Euclidean spaces and Lorentzian holonomy groups. Rend. Circ. Mat. Palermo. **79**, 87–97 (2006)
76. Leistner, T.: Towards a classification of Lorentzian holonomy groups. J. Diff. Geom. **76**, 423–484 (2007). [arXiv:math/0305139](#)
77. Leistner, T.: Towards a classification of Lorentzian holonomy groups. Part II: semisimple, non-simple weak-Berger algebras. J. Differ. Geom. **76**, 423–484 (2007). [arXiv:math/0309274](#)
78. Brinkmann, H.W.: Einstein spaces which are mapped conformally on each other. Math. Ann. **94**, 119–145 (1925)
79. Cahen, M., Wallach, N.: Lorentzian symmetric spaces. Bull. AMS. **76**, 585–591 (1970)
80. Penrose, R.: Any space-time has a plane wave as a limit. In: Differential Geometry and Relativity, pp. 271–275. Reidel, Dordrecht (1976)
81. Güven, R.: Plane wave limits and T-duality. Phys. Lett. B. **482**, 255–263 (2000). [arXiv:hep-th/0005061](#)
82. Blau, M., Figueroa-O'Farrill, J.M., Hull, C., Papadopoulos, G.: Penrose limits and maximal supersymmetry. Class. Quant. Grav. **19**, L87–L95 (2002). [arXiv:hep-th/0201081](#)

83. Blau, M., Figueroa-O'Farrill, J.M., Papadopoulos, G.: Penrose limits, supergravity and brane dynamics. *Class. Quant. Grav.* **19**, 4753 (2002). [arXiv:hep-th/0202111](#)
84. Blau, M., Borunda, M., O'Loughlin, M., Papadopoulos, G.: Penrose limits and space-time singularities. *Class. Quant. Grav.* **21**, L43 (2004). [arXiv:hep-th/0312029](#)
85. Maldacena, J.M., Maoz, L.: Strings on pp waves and massive two-dimensional field theories. *JHEP*. **12**, 046 (2002). [arXiv:hep-th/0207284](#)
86. Metsaev, R.R.: Type IIB Green-Schwarz superstring in plane wave Ramond-Ramond background. *Nucl. Phys. B*. **625**, 70 (2002). [arXiv:hep-th/0112044](#)
87. Figueroa-O'Farrill, J.M., Papadopoulos, G.: Homogeneous fluxes, branes and a maximally supersymmetric solution of M-theory. *J. High Energy Phys.* **06**, 036 (2001). [arXiv:hep-th/0105308](#)
88. Kowalski-Glikman, J.: Vacuum states in supersymmetric Kaluza-Klein theory. *Phys. Lett.* **134B**, 194–196 (1984)
89. Blau, M., Figueroa-O'Farrill, J.M., Hull, C.M., Papadopoulos, G.: A new maximally supersymmetric background of type IIB superstring theory. [arXiv:hep-th/0110242](#)
90. Berkovits, N., Maldacena, J.: $N = 2$ superconformal description of superstring in Ramond-Ramond plane wave backgrounds. *JHEP*. **0210**, 059 (2002). [arXiv:hep-th/0208092](#)
91. Russo, J.G., Tseytlin, A.A.: [arXiv:hep-th/0208114](#)

Chapter 9

Anomalies and All That



Abstract Quantum consistent string theories should be anomaly-free. In this chapter, we check this prediction. The cancelation of anomalies in the 16-SUSY strings is due to a novel mechanism (the *Green–Schwarz mechanism*) which also puts a topological constraint on the backgrounds where the string can propagate. We show that the absence of spacetime anomalies is an automatic consequence of modular invariance, i.e. absence of 2d gravitational anomalies in the world-sheet (perturbative) formulation. As a preliminary, we review the anomaly polynomial formalism for local anomalies and the Atiyah–Singer index theorems.

The presence of anomalies means that some classical symmetry is not preserved at the quantum level. Anomalies in local (i.e. gauged) symmetries make the theory inconsistent, since unphysical degrees of freedom no longer decouple. Our task is to study potential anomalies in target-space gauge and Diff^0 invariances. The string theories we constructed in Chaps. 5 and 7 are consistent at the quantum level, hence all local anomalies must cancel; in this chapter, we check *explicitly* that they do. The anomaly cancelation may be understood at two levels: one can check it in the full string theory, or just in the low-energy effective SUGRA. The two procedures are equivalent because the anomalies are fully captured by the massless sector of the theory [1]. Given the relevance of the issue, we shall consider anomaly cancelation from both points of view: first at the level of the effective SUGRA and then in the fully-fledged string theory. We start with a review of anomalies in QFT.

9.1 Review of the Anomaly Polynomial Formalism

General references for this section are [2–10].

We assume the spacetime X to have *even* dimension, $d \equiv 2n$, and positive (Euclidean) signature. We see the spacetime Riemann tensor as a 2-form $R^{(2)}$ with coefficients in the holonomy Lie algebra $\mathfrak{so}(d)$ (or $\mathfrak{so}(d-1, 1)$ in Lorentzian signature) acting in the vector representation: writing the indices explicitly, $(R^{(2)})_{ab} \equiv \frac{1}{2} R_{\mu\nu ab} dx^\mu \wedge dx^\nu$. Likewise, the gauge field strength is seen as a 2-form $F^{(2)} \equiv \frac{1}{2} F_{\mu\nu}^a dx^\mu \wedge dx^\nu$ with values in the gauge Lie algebra $\mathfrak{g}$ with generators t^a . We

combine the gauge and spin connections, $A^{(1)}$ and $\omega^{(1)}$, in a $\mathfrak{g} \oplus \mathfrak{so}(d)$ connection $\mathcal{A}^{(1)} = A^{(1)} + \omega^{(1)}$ with curvature $\mathcal{F}^{(2)} = R^{(2)} + F^{(2)}$.

The gravitational anomalies may be equivalently seen¹ either as **Diff**⁰ anomalies or as anomalies of the local $SO(d)$ symmetry which rotates the local frame (vielbein) e_μ^a . We take the second viewpoint. Then the full set of gravitational, gauge, and mixed gravitational-gauge anomalies have the formal structure of a gauge anomaly for the gauge group

$$\mathcal{G} \stackrel{\text{def}}{=} G \times SO(d), \quad (9.1)$$

with gauge connection $\mathcal{A}^{(1)}$ and curvature $\mathcal{F}^{(2)}$.

The theory of local² anomalies in QFT consist of two main issues:

- (i) the classification of the *possible* gauge/gravitational/mixed anomalies which may exist in a d -dimensional theory with gauge group G ; they form a finite-dimensional vector³ space $\mathbf{A}_{d,G}$;
- (ii) the determination of the vector in $\mathbf{A}_{d,G}$ which yields the actual anomalies for each given theory. The vector vanishes iff the QFT is anomaly-free.

We recall that a polynomial $I: \otimes^k \mathfrak{a} \rightarrow \mathbb{R}$ on the Lie algebra $\mathfrak{a}$ of a Lie group L , homogeneous of degree k in the elements $a_i \in \mathfrak{a}$, is **ad**-invariant symmetric iff

$$\begin{aligned} I(a_{\pi(1)}, \dots, a_{\pi(k)}) &= I(a_1, \dots, a_k), \quad \text{for all } \pi \in \mathfrak{S}_k \\ I(g^{-1}a_1g, \dots, g^{-1}a_kg) &= I(a_1, \dots, a_k) \quad \text{for all } g \in L. \end{aligned} \quad (9.2)$$

The answer to (i) is:

Claim 9.1 *The space $\mathbf{A}_{d,G}$ of possible local anomalies in $d \equiv 2n$ dimensions with gauge group G is the vector space of **ad**-invariant (symmetric) homogeneous polynomials on the Lie algebra $\mathfrak{g} \oplus \mathfrak{so}(d)$ of degree $n + 1$.*

The **ad**-invariant polynomial which yields the anomalies of a given QFT is called the *anomaly polynomial* of the theory.

It remains to justify the **Cla**. We write Θ for a parameter space with a base point 0, and fix a family of local gauge and frame rotations $g: X \times \Theta \rightarrow \mathcal{G}$ such that $g|_{X \times \{0\}}$ is the identity. We quantize our QFT in the spacetime X endowed with a family of gauge and gravitational backgrounds parametrized by Θ : the background at $\theta \in \Theta$ is obtained from the reference one at 0 by the gauge transformation $g(x; \theta) \in \mathcal{G}$. In the absence of anomalies, the gauge transformation is a quantum symmetry, and the partition function $Z(\theta)$ is constant in Θ . By definition, the transformations are anomalous iff $Z(\theta)$ is not constant; more precisely its absolute value $|Z(\theta)|$ is still constant, but its phase now does depend on $\theta \in \Theta$.

¹ For a proof of this statement see Sect. 5 of [8].

² *Local* as contrasted to *global*; in this section, we are only concerned with quantum anomalies of classical symmetries given by field transformations which are homotopic to the identity.

³ Actually they form a lattice $\mathbf{A}_{d,G}^{\mathbb{Z}} \subset \mathbf{A}_{d,G}$ in this vector space. We ignore the integral structure for the moment; it will be clear from the relations of anomalies with characteristic classes.

We say that a differential form on $X \times \Theta$ of the form $C dx^{i_1} \cdots dx^{i_k} d\theta^{a_1} \cdots d\theta^{a_l}$ has *bi-degree* (k, l) . We write $\Lambda^{k,l}$ for the space of forms of given bi-degree, and $d: \Lambda^{k,l} \rightarrow \Lambda^{k+1,l}$, $\delta: \Lambda^{k,l} \rightarrow \Lambda^{k,l+1}$ for the de Rham differentials acting on forms on X and Θ , respectively. The total de Rham differential in $X \times \Theta$ is

$$d = d + \delta: \Lambda^{k,l} \rightarrow \Lambda^{k+1,l} \oplus \Lambda^{k,l+1}. \quad (9.3)$$

We set $v \equiv g^{-1}\delta g$: it is a $(0, 1)$ -form with coefficients in the Lie algebra $\mathfrak{g} \oplus \mathfrak{so}(d)$. v , being anticommuting, may be identified with the Fadeev–Popov ghost; then δ is the BRST operator.

When our theory is plagued by anomalies, the variation of the effective action $\log Z$ along Θ is not zero, but the integral of a local operator of dimension $d \equiv 2n$

$$\delta \log Z = \frac{i}{(2\pi)^n} \int_X \Omega(x)^{(d,1)}. \quad (9.4)$$

$\Omega(x)^{(d,1)}$ is a d -form in X which is a local polynomial in the background fields and their derivatives. $\Omega(x)^{(d,1)}$ is linear in v , hence of bi-degree $(d, 1)$. The d -form $\Omega^{(d,1)}$ in Eq. (9.4) is our anomaly. Our task is to classify all possible operators $\Omega^{(d,1)}$ which may play the role of anomaly in d dimension and gauge group $\mathcal{G}$. They are restricted by a consistency requirement: the *Wess-Zumino condition* [11]; since $\delta^2 = 0$, we must have

$$\delta \int_X \Omega^{(d,1)} = 0. \quad (9.5)$$

The anomaly $\Omega^{(d,1)}$ is not unique on two counts: (1) we may add to it a total derivative $d\xi^{(d-1,1)}$ which integrates to zero, and (2) we are free to add to the effective action $\log Z$ any local counter-term $\int_X \mathcal{O}(x)^{(d,0)}$; this has the effect of shifting

$$\Omega^{(d,1)} \rightarrow \Omega^{(d,1)} + \delta \mathcal{O}^{(d,0)} \quad (9.6)$$

Hence two operators, $\Omega^{(d,1)}$ and $\hat{\Omega}^{(d,1)}$, which differ by $\delta \mathcal{O}^{(d,0)} + d\xi^{(d-1,1)}$ give equivalent anomalies. Thus the physically distinct anomalies correspond to δ -closed local functionals modulo the δ -exact ones, i.e. the space of possible anomalies $\mathbf{A}_{2n,\mathcal{G}}$ is the *cohomology* of the differential δ computed in the space⁴ of local functionals constructed out of the connections and their derivatives.

Anomaly Descent

Claim 9.1 says that the δ -cohomology on local functionals of the connection is given by the **ad**-invariant symmetric polynomials $I: \odot^{n+1}(\mathfrak{g} \oplus \mathfrak{so}(2n)) \rightarrow \mathbb{R}$,

⁴ As it will be clear below, it is also the cohomology of local operators for the total differential d . Indeed, $d\xi = (d + \delta)\xi$ is the sum of a δ -exact form and a d -exact one.

homogeneous of degree $n + 1$. We see an **ad**-invariant polynomial I as a *formal* $(d + 2)$ -form by evaluating it⁵ on the curvature 2-form $\mathcal{F}^{(2)} \equiv F^{(2)} + R^{(2)}$:

$$I^{(d+2)}(R^{(2)}, F^{(2)}) \equiv I(\mathcal{F}^{(2)}, \dots, \mathcal{F}^{(2)}). \quad (9.7)$$

$I^{(d+2)}(R^{(2)}, F^{(2)})$ can be written as a sum of products of “single trace” invariants of the form $\text{tr}[P(R^{(2)}, F^{(2)})]$ where $P(\cdot, \cdot)$ is a polynomial and the trace is over the holonomy and matter gauge representations. More intrinsically, by the Harish–Chandra isomorphism⁶ the **ad**-invariant polynomials on any Lie algebra $\mathcal{L}$ form a free polynomial algebra with $r \equiv \text{rank}(\mathcal{L})$ generators whose degrees d_i are (by definition) the degrees of the rational reflection group $\text{Weyl}(\mathcal{L})$; one has $d_i = \ell_i + 1$ with ℓ_i the exponents of $\text{Weyl}(\mathcal{L})$. By the Bianchi identities, we have

$$dI^{(d+2)}(\mathcal{F}^{(2)}) = (n + 1) I(\mathcal{F}^{(2)}, \dots, \mathcal{F}^{(2)}, D\mathcal{F}^{(2)}) = 0, \quad (9.8)$$

and hence $I^{(d+2)}(\mathcal{F}^{(2)})$ may be written as the exterior derivative of a $(d + 1)$ -form $\omega^{(d+1)}(\mathcal{A}^{(1)}, \mathcal{F}^{(2)})$ which is *not* globally defined since $I^{(d+2)}(R^{(2)}, F^{(2)})$ represents a non-trivial *characteristic class* of the underlying $G \times SO(2n)$ principal bundle [14–16]. The $(d + 1)$ -form $\omega^{(d+1)}(\mathcal{A}^{(1)}, \mathcal{F}^{(2)})$ is called the *Chern–Simons form* associated to the **ad**-invariant polynomial I . The explicit form of $\omega^{(d+1)}(\mathcal{A}^{(1)}, \mathcal{F}^{(2)})$ for a given **ad**-invariant polynomial I is given, e.g. in the appendix of [17]; clearly it is a local operator constructed with the connections and their derivatives.

We consider the gauge connection on the “big” space $X \times \Theta$

$$\mathcal{A} = g^{-1}\mathcal{A}^{(1)}g + g^{-1}dg = \mathcal{A}_g^{(1)} + v, \quad (9.9)$$

where $\mathcal{A}_g^{(1)} \equiv g^{-1}\mathcal{A}^{(1)}g + g^{-1}dg$ is the usual gauge transform of $\mathcal{A}^{(1)}$. We set

$$\mathcal{F} \equiv d\mathcal{A} + \mathcal{A}^2 = g^{-1}\mathcal{F}^{(2)}g. \quad (9.10)$$

The argument following Eq. (9.8) applies both to X and $X \times \Theta$, so

$$\begin{aligned} d\omega^{(d+1)}(\mathcal{A}^{(1)}, \mathcal{F}^{(2)}) &= I^{(d+2)}(\mathcal{F}^{(2)}) = I^{(d+2)}(\mathcal{F}) = \\ &= d\omega^{(d+1)}(\mathcal{A}, \mathcal{F}) \equiv (d + \delta)\omega^{(d+1)}(\mathcal{A}_g^{(1)} + v, \mathcal{F}), \end{aligned} \quad (9.11)$$

where we used (9.10) and the fact that $I^{(d+2)}$ is **ad**-invariant. We expand the $(d + 1)$ -form $\omega^{(d+1)}(\mathcal{A}_g^{(1)} + v, \mathcal{F})$ in forms of definite bi-degree

$$\omega^{(d+1)}(\mathcal{A}_g^{(1)} + v, \mathcal{F}) = \sum_l \omega^{(d+1-l, l)}, \quad (9.12)$$

⁵ The curvature 2-forms, having even degree, generate a commutative algebra, hence a quotient of a polynomial algebra by an ideal. So it makes sense to evaluate polynomials on them.

⁶ See, e.g. [12] Chap. VIII Sect. 8 or Sect. V.5 of [13].

where $\omega^{(d+1-l,l)}$ is of order l in v and its derivatives. Expanding in bi-degrees, the identity (9.11), we get the chain of identities (called the *descent equations*)

$$\delta\omega^{(d+1-l,l)} = -d\omega^{(d-l,l+1)} \quad \text{for } l = 0, 1, 2, \dots, d+1. \quad (9.13)$$

In particular, the d -form

$$\Omega^{(d)} \equiv \omega^{(d,1)}(v, \mathcal{A}^{(1)}, \mathcal{F}^{(2)}) \quad (9.14)$$

satisfies the Wess–Zumino condition. $\Omega^{(d)}$ is the anomaly associated to the **ad**-invariant polynomial $I^{(d+2)}(\mathcal{F}^{(2)})$ implied by the **Claim**. That this map between **ad**-invariant polynomials $I^{(d+2)}$ and potential anomalies $\Omega^{(d)}$ is one-to-one follows from the relation of anomalies in d dimensions with the index theorem in $d+2$ dimensions we are going to discuss.

Relation with the Atiyah–Singer Index Theorem

The anomaly polynomial $I^{(d+2)}$ for each given QFT is determined by the Lorentz and gauge representations of the light fields: $I^{(d+2)}$ is the sum of the anomaly polynomials associated to the various massless fields with given spin and gauge representation. Only chiral fields (fermionic and bosonic) contribute: their e.o.m. can be written in the Dirac form $\mathfrak{D}\psi = \dots$ where ψ is a section of a spin bundle of definite chirality $S_{\pm}$ twisted by some $G \times \text{Spin}(d)$ vector bundle V which depends on the massless field: the differential operator $\mathfrak{D}$ is an ordinary Dirac operator acting on Weyl fermions coupled to the natural connection on the $G \times \text{Spin}(d)$ bundle V .

Example: $\mathfrak{D}$ for chiral bosons in IIB

The full set of R-R massless fields in IIB are described in the natural $(-\frac{1}{2}, -\frac{3}{2})$ picture by an even-degree gauge-form $A_e = \sum_k A^{(2k)}$ which we see as a bispinor $A_{\alpha\dot{\rho}}$; the bispinor is chiral iff the corresponding form A_e is an eigenform of the Hodge operator $*$; one sees $A_{\alpha\dot{\rho}}$ as a Weyl spinor in the index α twisted by the vector bundle S_- (the negative spin bundle) associated to the “extra” index $\dot{\rho}$. The operator $\mathfrak{D}$ acts on $A_{\alpha\dot{\rho}}$ as the usual Weyl Dirac operator, $i(\Gamma_{\mu})^{\dot{\rho}\alpha} D^{\mu}$, where the Γ -matrices act only on the first index α and D_{μ} contains, besides the usual S_+ spin connection, a connection on the twisting bundle S_- . This Dirac operator $\mathfrak{D}$ is precisely the one predicted by the BRST condition in Chap. 3 which reads $\mathfrak{D}A_e = 0$. For its index, see **BOX 9.2**.

The answer to (ii) is [8]:

Claim 9.2 *The anomaly polynomial $I^{(d+2)}$ for a chiral field is essentially the Atiyah–Singer index density⁷ for the corresponding differential operator $\mathfrak{D}$ in $(d+2)$ dimensions. Example for chiral spin- $\frac{1}{2}$ field $I^{(d+2)}$ is precisely the index density⁸*

$$\text{index } \mathfrak{D} = \frac{1}{(2\pi)^{n+1}} \int I^{(d+2)} \Big|_{\text{spin}} \frac{1}{2} \quad \text{in } d+2 \text{ dimensions.} \quad (9.15)$$

⁷ For the Atiyah–Singer index theorem see [18–25].

⁸ The numerical normalizations are chosen for later convenience.

An argument for this equality is sketched in **BOX 9.1**. The details of the relation between the index and the anomaly polynomial in the other cases, in particular, for the self-dual forms, are discussed in **BOX 9.2**.

BOX 9.1 - The anomaly polynomial is the index density in $d + 2$

It suffices to consider the gauge anomaly. Let A be our gauge connection on the d -dimensional space X ; we may assume that the Weyl operator $\mathfrak{D} \equiv \frac{i}{2}(1 + \gamma_{d+1})\gamma^\mu(\partial_\mu - iA_\mu)$ has no zero-modes. Following [8], we consider the $(d + 2)$ -dimensional space $X \times S^2$ with a gauge field $\mathcal{A} \equiv \mathcal{A}_\mu dx^\mu$ of the following form: on the lower hemisphere $\mathcal{A} \equiv A$, while on the upper hemisphere we have a smooth connection $\mathcal{A} \equiv \mathcal{A}'$, and the two hemispheres are glued along the equator $|z| = 1$ by a gauge transformation $g(\theta)$, i.e. $\mathcal{A}'|_{z=e^{i\theta}} = g(\theta)^{-1}A g(\theta) + g(\theta)^{-1}dg(\theta)$. The net number of the chiral zero-modes in $(d + 2)$ dimensions is given by the index theorem

$$\text{index}\left\{\frac{i}{2}(1 + \Gamma_{d+3})\Gamma^M(\partial_M - i\mathcal{A}_M)\right\} = \int_{X \times S^2} \frac{I^{(d+2)}}{(2\pi)^{n+1}} = \int_{X \times S^1} \frac{\omega_{\text{upper}}^{(d+1)} - \omega_{\text{lower}}^{(d+1)}}{(2\pi)^{n+1}}$$

Only the term linear in $g^{-1}\partial_\theta g$, i.e. of bi-degree $(d, 1)$, gives a non-zero contribution, that is,

$$\#(\text{chiral zero-modes}) = \int_0^{2\pi} d\theta \int_X (2\pi)^{-(n+1)} \omega^{(d,1)}|_{v \equiv g^{-1}\partial_\theta g} \quad (\spadesuit)$$

The index is invariant under continuous deformations; we may replace $\Gamma^M D_M \equiv \gamma^\mu D_\mu + \gamma^i D_i$ with $\lambda \mathcal{D} + \gamma^i D_i$ and take $\lambda \rightarrow \infty$. In this limit the zero-modes concentrate around points in the North hemisphere where the d -dimensional Dirac operator $\mathcal{D}$ has a zero mode. The index counts the number of such points in the upper hemisphere (with multiplicity).

Now compute the path integral $Z \equiv \det \mathcal{D}$ of the d -dimensional spinor as a function of θ along the equator. By definition

$$\partial_\theta \log Z(\theta) = i \int_X (2\pi)^{-(n+1)} \Omega^{(d)}|_{v \equiv g^{-1}\partial_\theta g}$$

so that the total winding number of the phase of $Z(\theta)$ as θ goes from 0 to 2π is given by the anomaly

$$\text{winding number of } \arg Z = \int_0^{2\pi} \frac{d\theta}{2\pi} \int_X (2\pi)^{-n} \Omega^{(d)}|_{v \equiv g^{-1}\partial_\theta g} \quad (\clubsuit)$$

But the winding number is also the number of zeros of the function $Z(z)$ on the upper hemisphere $|z| < 1$ (counted with multiplicities); by the previous argument this number is the index. Comparing Eqs. $(\spadesuit)$ and $(\clubsuit)$, we get the desired equality $\Omega^{(d)} \equiv \omega^{(d,1)}$.

For Weyl (i.e. *complex*) fermions and bispinors, we normalize the anomaly polynomials as follows (to avoid cluttering, we replace $R^{(2)}$ by R and $F^{(2)}$ by F):

$$\frac{1}{(2\pi)^{n+1}} I^{(d+2)} \Big|_{\text{spin } \frac{1}{2}} = \left[\widehat{A}(R) \text{Ch}_\varrho(F) \right]^{(d+2)} \quad (9.16)$$

$$\frac{1}{(2\pi)^{n+1}} I^{(d+2)} \Big|_{\text{spin } \frac{3}{2}} = \left[\widehat{A}(R) (\text{Ch}(R) - 1) \right]^{(d+2)} \quad (9.17)$$

$$\frac{1}{(2\pi)^{n+1}} I^{(d+2)} \Big|_{\text{self-dual}} = -\frac{1}{4} \left[L(R) \right]^{(d+2)} \quad (9.18)$$

where $[\dots]^{(k)}$ means taking the k -form part of the expression inside the bracket, and the RHS is the Atiyah–Singer index in $d + 2$ dimensions for, respectively, a Weyl spin-1/2, a Weyl spin-3/2,⁹ and a complex bispinor; the spin-1/2 transforms in the representation ϱ of the gauge group. Here [18–25, 27]:

- $\widehat{A}(R)$ is the *Dirac genus* of the spacetime manifold M ;
- $\text{Ch}_\varrho(F)$ is the *Chern character* of the gauge vector bundle $V_\varrho \rightarrow M$ associated to the G -representation ϱ ;
- $L(R)$ is the *Hirzebruch L -polynomial* of the spacetime manifold M .

For Majorana–Weyl fermions and real bispinors, we have to take 1/2 of the above expressions. We stress that the self-dual form anomaly has an extra minus sign, since the statistics is now bosonic (instead of fermionic) and hence the one-loop path integral is an inverse determinant rather than a determinant.

Exercise 9.1 Show that, as predicted by the above **Example**, the index for a complex chiral bosons is minus the index for a Weyl fermion coupled to a spin bundle.

Formulae for the Characteristic Classes If x_a are the skew-eigenvalues of the curvature 2-form $R^a{}_b/2\pi$, the index of a Dirac operator coupled to a connection with curvature F is¹⁰ [18–24]

$$\text{ind } i\mathcal{D} = \int \widehat{A}(R) \text{Ch}_\varrho(F), \quad (9.19)$$

$$\text{Ch}_\varrho(F) = \text{tr}_\varrho e^{iF/2\pi}, \quad \widehat{A}(R) = \prod_a \frac{x_a/2}{\sin(x_a/2)}. \quad (9.20)$$

The signature τ of a (compact, oriented) manifold M is given by [27]

⁹ As we saw in Sect. 8.1, the massless gravitini are the gauge fields of local supersymmetry; thus to quantize them, we need to fix the SUSY gauge and introduce the corresponding (commuting) Fadeev–Popov ghost: Eq. (9.17) is the resulting total index with the contribution from the (chiral) spinorial ghosts *included*. For more details, see [26].

¹⁰ Let $f(x)$ be a power series starting with 1 as $(x/2)/\sin(x/2)$ in (9.20) or $x/\tanh x$ in (9.21). The precise meaning of the formal expression $\prod_a f(x_a)$ is as follows: write $\prod_a f(x_a) = \sum_k p_k(x_a)$ where $p_k(x_a)$ are degree- k homogeneous polynomials symmetric in the x_a which can be rewritten as polynomials $q_k(h_\ell)$ in the degree- ℓ symmetric polynomials $h_\ell \equiv \sum_a x_a^\ell = \text{tr}(R/2\pi)^\ell$. Then $\prod_a f(x_a) = \sum_k q_k(\text{tr}(R/2\pi)^\ell)$, whose components of form-degree $2k$, $q_k(\text{tr}(R/2\pi)^\ell)$, are well-defined **ad**-invariant symmetric polynomials in the curvature 2-form R .

$$\tau = \int_M L(R), \quad \text{where} \quad L(R) = \prod_a \frac{x_a}{\tanh x_a}. \quad (9.21)$$

τ vanishes unless $\dim M \equiv 4m$ is a multiple of 4; in this case, $\tau = b_{2m}^+ - b_{2m}^-$ where b_{2m}^+ (resp. b_{2m}^-) is the dimension of the space of self-dual (resp. anti-self-dual) harmonic $2m$ -forms. Explicitly, the first few terms in the degree expansions read

$$\begin{aligned} \widehat{A}(R) = & 1 + \frac{1}{(4\pi)^2} \frac{1}{12} \text{tr} R^2 + \frac{1}{(4\pi)^4} \left[\frac{1}{288} (\text{tr} R^2)^2 + \frac{1}{360} \text{tr} R^4 \right] + \\ & + \frac{1}{(4\pi)^6} \left[\frac{1}{10368} (\text{tr} R^2)^3 + \frac{1}{4320} \text{tr} R^2 \text{tr} R^4 + \frac{1}{5670} \text{tr} R^6 \right] + O(R^8) \end{aligned} \quad (9.22)$$

$$\begin{aligned} L(R) = & 1 - \frac{1}{(2\pi)^2} \frac{1}{6} \text{tr} R^2 + \frac{1}{(2\pi)^4} \left[\frac{1}{72} (\text{tr} R^2)^2 - \frac{7}{180} \text{tr} R^4 \right] + \\ & + \frac{1}{(2\pi)^6} \left[-\frac{1}{1296} (\text{tr} R^2)^3 + \frac{7}{1080} \text{tr} R^2 \text{tr} R^4 - \frac{31}{2835} \text{tr} R^6 \right] + O(R^8). \end{aligned} \quad (9.23)$$

In the index density for the spin- $\frac{3}{2}$ RS operator, Eq. (9.17),

$$\text{Ch}(R) \equiv \text{tr}(e^{iR/2\pi} - 1) + d \quad (9.24)$$

since R is a $SO(d)$ curvature not a $(d+2)$ -dimensional one [8].

Specializing to $d = 10$ In 10d the anomaly 12-form are (here $n = \dim \varrho$)

$$\begin{aligned} \frac{1}{2} I_{\text{spin } 1/2}^{(12)} = & -\frac{1}{1440} \text{tr}_\varrho F^6 + \frac{1}{2304} \text{tr}_\varrho F^4 \text{tr} R^2 - \frac{1}{23040} \text{tr}_\varrho F^2 \text{tr} R^4 - \frac{1}{18432} \text{tr}_\varrho F^2 (\text{tr} R^2)^2 \\ & + \frac{n}{725760} \text{tr} R^6 + \frac{n}{552960} \text{tr} R^4 \text{tr} R^2 + \frac{n}{1327104} (\text{tr} R^2)^3 \end{aligned} \quad (9.25)$$

$$\frac{1}{2} I_{\text{spin } 3/2}^{(12)} = -\frac{495}{725760} \text{tr} R^6 + \frac{225}{552960} \text{tr} R^4 \text{tr} R^2 - \frac{63}{1327104} (\text{tr} R^2)^3 \quad (9.26)$$

$$\frac{1}{2} I_{\text{self-dual}}^{(12)} = \frac{992}{725760} \text{tr} R^6 - \frac{448}{552960} \text{tr} R^4 \text{tr} R^2 + \frac{128}{1327104} (\text{tr} R^2)^3. \quad (9.27)$$

By definition, the terms in the polynomials which contain only F (resp. R) give the pure gauge (resp. gravitational) anomalies. The terms which contain both curvatures yield the mixed gauge-gravitational anomalies.

9.2 Anomaly Cancellation in 10d SUSY String Theories

The IIA theory is parity-symmetric in spacetime, hence automatically anomaly-free. Let us consider the anomalies in the *chiral* 10d string theories.

Anomaly Cancellation in Type IIB

In Type IIB, there are no gauge fields, so no gauge or mixed anomalies. The chiral massless sector of Type IIB consists of two Majorana–Weyl gravitini of the same chirality, two Majorana–Weyl dilatini with chirality opposite to the gravitini, and one *real* self-dual field strength. Hence the total Type IIB anomaly polynomial is

BOX 9.2 - Indices for the Kähler–Dirac operator $d + \delta$

We discuss the indices for the operator $d + \delta$; cf. Sect. 25.4. (b) of [27]. The space of all differential forms $\Omega^\bullet$ decomposes in the spaces of even/odd forms $\Omega^\bullet = \Omega_{ev} \oplus \Omega_{od}$. On a Euclidean $4k$ -fold, we define the operator α which acts on p -forms as $i^{p(p+1)-2k} *$. Then $\alpha^2 = 1$ and one decomposes the space of all forms according to the ± 1 eigenvalue of α , $\Omega^\bullet = \Omega_+^\bullet \oplus \Omega_-^\bullet$. One has

$$\alpha(d + \delta) = -(d + \delta)\alpha.$$

We may define 4 different indices depending on the domain/range of the operator $d + \delta$

I_e	I_+	$I_{e,+}$	$I_{e,-}$
$d + \delta: \Omega_{ev} \rightarrow \Omega_{od}$	$d + \delta: \Omega_+ \rightarrow \Omega_-$	$d + \delta: \Omega_{ev,+} \rightarrow \Omega_{od,-}$	$d + \delta: \Omega_{ev,-} \rightarrow \Omega_{od,+}$

We have $I_+ = I_{e,+} - I_{e,-}$ and $I_e = I_{e,+} + I_{e,-}$. Moreover, by definition,

$$I_e = \chi \quad \text{Euler characteristic} \qquad I_+ = \tau \equiv \int L(R) \quad \text{Hirzebruch signature}$$

Subtleties in the Anomaly Polynomial for Self-dual Tensors The anomaly polynomial for the self-dual field strength (and for the gravitino) is not exactly the index density in two more dimensions. In order to see this, let us return to the string description of the self-dual form as a RR field in type IIB, cf. **BOX 3.3**. We know that the natural geometrical description of a RR vertex is in picture $(-\frac{1}{2}, -\frac{3}{2})$ where the RR fields are described in terms of a covariant bispinor of the form $\phi_{\alpha\beta}$ whose two indices have opposite chirality: such a bispinor is an *anti-self-dual odd* form. The BRST condition yields a Dirac operator acting on the first index $\not{D}\phi = 0$, but there is no Dirac condition on the second index (as one would have in the $(-\frac{1}{2}, -\frac{1}{2})$ picture). Hence $\phi_{\alpha\beta}$ should be seen as a positive chirality Weyl spinor carrying an extra index $\hat{\beta}$ which couples to the spin connection of spacetime. The anomaly polynomial of such a bispinor is not given by the index in $(d+2)$ dimensions of a bispinor, but rather by the index in $(d+2)$ dimensions of a spinor coupled to the d -dimensional spin connection (instead of the $(d+2)$ -dimensional one). Denoting by a hat the quantities so defined, we have $\hat{I}_e = 0$ (no anomaly in the parity invariant combination) and $\hat{I}_+ = I_+/2$ [8]. Hence

$$\hat{I}_{+,e} = \frac{1}{4} I_+ \equiv \frac{1}{4} \tau \equiv \frac{1}{4} \int L(R).$$

This is for a *complex* self-dual field: in the *real* case we have an additional factor $1/2$.

$$\begin{aligned}
I_{\text{IIB}}^{(12)} &= -I_{\text{spin } 1/2}^{(12)} + I_{\text{spin } 3/2}^{(12)} + \frac{1}{2} I_{\text{self-dual}}^{(12)} = \\
&= \frac{-2 - 2 \times 495 + 992}{725760} \text{tr } R^6 + \frac{-2 + 2 \times 225 - 448}{552960} \text{tr } R^4 \text{tr } R^2 + \\
&\quad + \frac{-2 - 2 \times 63 + 128}{1327104} (\text{tr } R^2)^3 = 0.
\end{aligned} \tag{9.28}$$

The intricate numerical coefficients in Eqs. (9.25)–(9.27) magically conspire so that the total anomaly vanishes. From the viewpoint of the low-energy theory, this looks an unexpected miracle, but we know that the theory is consistent at the quantum level (since it satisfies the stringy consistency conditions) and so it should be, in particular, non-anomalous. In Sect. 9.3, we shall understand the stringy origin of the anomaly cancelation in a more direct and conceptual way.

Type I and Heterotic Anomalies: The Green–Schwarz Mechanism

Type I and $SO(32)$ heterotic strings have the same low-energy theory, hence the same anomalies. At first sight, it seems that here we encounter an insurmountable obstacle: the *only* chiral fields charged under the gauge group G are the gluini, so apparently there is no possibility to cancel their gauge and mixed anomalies with contributions coming from other fields, as we did in Type IIB. This is puzzling since we proved in Chaps. 5 and 7 that they are quantum consistent, hence anomaly-free. The paradox was solved by Green and Schwarz who demonstrated how the anomaly is canceled in a different way which is now called the *Green–Schwarz mechanism* after them [28]. The point is that our assertion in Sect. 9.1 that the chiral anomalies constructed from the index density in two more spacetime dimensions are *cohomologically non-trivial*, that is, cannot be canceled by a *local* counter-term, was based on the assumption that all allowed counter-terms are local in the gauge and metric fields. In string theory, we have additional light fields, and hence a larger zoo of possible counter-terms which we may use to cancel some potential anomaly that would be unavoidable in the usual field-theoretic set-up.

Gauge Anomalies To illustrate the Green–Schwarz mechanism in a simple situation, let us consider the Chern–Simons interaction¹¹

$$S' = \int B^{(2)} \wedge \text{tr } F^4 \tag{9.29}$$

where the trace is taken in some representation ϱ of the gauge group G . The coupling (9.29) is invariant under the 2-form gauge transformation

$$\delta B^{(2)} = d\xi^{(1)} \tag{9.30}$$

by integration by parts and the Bianchi identity which implies $d \text{tr } F^4 = 0$. We know from Sect. 8.6 that in 10d $\mathcal{N} = 1$ supergravity the field $B^{(2)}$ has the non-trivial gauge transformation (cf. Eq. (8.104))

¹¹ To avoid cluttering, in this paragraph, we write the YM field strength simply F instead of $F^{(2)}$.

$$\delta B^{(2)} \propto \text{tr}(vF), \quad (9.31)$$

under the ordinary infinitesimal Yang-Mills gauge transformation,

$$\delta A^{(1)} = Dv, \quad (9.32)$$

and so

$$\delta S' \propto \int \text{tr}(vF) \text{tr} F^4, \quad (9.33)$$

which has the general form of the anomaly in Eqs. (9.8)–(9.14) for the anomaly polynomial $I^{(12)} = \text{tr} F^2 \text{tr} F^4$ whose descendent forms $\omega^{(11-k,k)}$ may be chosen to be

$$\omega^{(11,0)} = \text{tr} \left(AdA + \frac{2}{3} A^3 \right) \text{tr} F^4 \quad (9.34)$$

$$\Omega^{(10,1)} \equiv \omega^{(10,1)} = \text{tr}(vF) \text{tr} F^4. \quad (9.35)$$

Thus the part of the anomaly associated to the monomial $\text{tr} F^2 \text{tr} F^4$ can be canceled by the local counter-term (9.29). Similarly, the variation of the local counter-term

$$S'' = \int B^{(2)} [\text{tr} F^2]^2 \quad (9.36)$$

can cancel the anomaly arising from a term in $I^{(12)}$ proportional to $[\text{tr} F^2]^3$.

However, the *pure* gauge anomaly polynomial is (cf. Eq. (9.25))

$$\frac{1}{2} I_{\text{spin } 1/2}^{(12)} \Big|_{R=0} = -\frac{1}{1440} \text{tr}_{\text{ad}} F^6, \quad (9.37)$$

where the trace is in the adjoint representation since the MW fermions are gluini. The above counter-terms S' , S'' cannot cancel single-trace anomalies of the form (9.37). Only one “miracle” may solve the theory; for each particular Lie algebra $\mathfrak{g}$, we have specific relations between its various **ad**-invariant polynomials, so *for special* $\mathfrak{g}$ it may be possible to rewrite $\text{tr}_{\text{ad}} F^6$ as a polynomial in traces of lesser powers of F , so that a suitable linear combination of the two counter-terms (9.29) and (9.36) may cancel all pure gauge anomalies.

Polynomial Relations Between $\text{tr}_{\text{ad}} F^{2k}$ for $SO(n)$ Gauge Groups When the gauge group is $SO(n)$ we write the **ad**-invariant polynomials in terms of traces in the vector

representation V . The adjoint representation of $SO(n)$ is $\wedge^2 V$. Elementary character theory¹² yields the identity

$$\mathrm{tr}_{\wedge^2 V}(e^{\lambda F}) = \frac{1}{2}(\mathrm{tr} e^{\lambda F})^2 - \frac{1}{2}\mathrm{tr} e^{2\lambda F}, \quad (9.38)$$

where we write tr for tr_V . Expanding (9.38) in powers of the parameter λ , we get for $\mathfrak{g} \equiv \mathfrak{so}(n)$ and $k \in \mathbb{N}$

$$\mathrm{tr}_{\mathrm{ad}} F^{2k} = \frac{1}{2} \sum_{\ell=0}^k \binom{2k}{2\ell} \mathrm{tr} F^{2\ell} \mathrm{tr} F^{2(k-\ell)} - 2^{2k-1} \mathrm{tr} F^{2k}. \quad (9.39)$$

In particular, this yields

$$\text{for } k = 1 \quad \mathrm{tr}_{\mathrm{ad}} F^2 = (n - 2) \mathrm{tr} F^2, \quad (9.40)$$

$$\text{for } k = 2 \quad \mathrm{tr}_{\mathrm{ad}} F^4 = (n - 8) \mathrm{tr} F^4 + 3(\mathrm{tr} F^2)^2, \quad (9.41)$$

$$\text{for } k = 3 \quad \mathrm{tr}_{\mathrm{ad}} F^6 = (n - 32) \mathrm{tr} F^6 + 15 \mathrm{tr} F^2 \mathrm{tr} F^4. \quad (9.42)$$

Equality (9.42) says that *precisely for $SO(32)$* the invariant polynomial $\mathrm{tr}_{\mathrm{ad}} F^6$ is equal to a product of lower power traces, and so the corresponding anomaly can be canceled by the variation of a linear combination of the two counter-terms S' and S'' . *This is the Green–Schwarz mechanism for anomaly cancelation* [28].

$SO(32)$ is the same gauge group which appears in Type I and heterotic string theories, that we already know to be quantum-consistent systems and hence anomaly-free. The above computation confirms this fact. Note that for $SO(32)$

$$\mathrm{tr} F^2 = \frac{1}{30} \mathrm{tr}_{\mathrm{ad}} F^2 = \frac{1}{h} \mathrm{tr}_{\mathrm{ad}} F^2, \quad \mathrm{tr} F^4 = \frac{1}{24} \mathrm{tr}_{\mathrm{ad}} F^4 - \frac{1}{8h^2} (\mathrm{tr}_{\mathrm{ad}} F^2)^2 \quad (9.43)$$

where $h \equiv 30$ is the Coxeter number of $SO(32)$. Moreover $16 \cdot h^2 = 14400$. Hence for $G = SO(32)$ we have the identity

$$\mathrm{tr}_{\mathrm{ad}} F^6 = \frac{1}{48} \mathrm{tr}_{\mathrm{ad}} F^2 \mathrm{tr}_{\mathrm{ad}} F^4 - \frac{1}{14400} (\mathrm{tr}_{\mathrm{ad}} F^2)^3. \quad (9.44)$$

To write uniform formulae for all $\mathfrak{g}$, it is convenient to define the normalized trace as $\mathrm{tr}(\cdots) = h^{-1} \mathrm{tr}_{\mathrm{ad}}(\cdots)$. For $\mathfrak{so}(n)$ this is the trace in the vector representation.

Relations Between ad-Invariant Polynomials for General Gauge Groups For a general Lie algebra $\mathfrak{g}$, the graded ring of **ad**-invariant polynomials, $I(F)$, is a polynomial ring in $r \equiv \mathrm{rank} \mathfrak{g}$ homogeneous generators of degrees equal to the exponents +1 of $\mathfrak{g}$. The degrees of E_8 are [30]

$$2, 8, 12, 14, 18, 20, 24, 30. \quad (9.45)$$

¹² Advanced readers may prefer to think in terms of λ -rings [29].

Hence in E_8 all traces of the form $\text{tr}_{\text{ad}} F^k$ with $k < 8$ may be written as polynomials in the trace $\text{tr}_{\text{ad}} F^2$. Then for E_8

$$\text{tr}_{\text{ad}} F^{2k} = c_k (\text{tr}_{\text{ad}} F^2)^k \quad \text{for } k \leq 3, \quad (9.46)$$

for certain numerical coefficients c_k ($c_k = 1$). To compute c_k consider the subgroups

$$SO(2) \subset SO(2) \times SO(14) \subset SO(16) \subset E_8 \quad (9.47)$$

and take F to be in the Lie algebra of the $SO(2)$ subgroup. Under $SO(16)$ the adjoint of E_8 decomposes in the adjoint plus a Majorana–Weyl spinor; hence under the subgroup $SO(2) \times SO(14) \simeq SO(14) \times U(1)$ it decomposes as

$$248 \rightsquigarrow \mathbf{91}_0 \oplus \mathbf{14}_{+1} \oplus \mathbf{14}_{-1} \oplus \mathbf{1}_0 \oplus \mathbf{64}_{+1/2} \oplus \mathbf{64}_{-1/2}. \quad (9.48)$$

Then, if F belongs to the $\mathfrak{so}(2) \simeq \mathfrak{u}(1)$ subalgebra, we have

$$\text{tr}_{\text{ad}} F^{2k} = \left(28 + \frac{128}{2^{2k}} \right) F^{2k} \quad (9.49)$$

which inserted in the identity (9.46) yields

$$\frac{1}{c_k} = \frac{(28 + 128/4)^k}{28 + 128 \cdot 2^{-2k}} \equiv \frac{240^k}{28 \cdot 2^{2k} + 128}. \quad (9.50)$$

So that

$$\frac{1}{c_2} = 100, \quad \frac{1}{c_3} = 7200, \quad \Rightarrow \quad c_3 \equiv \frac{c_2}{48} - \frac{1}{14400}, \quad (9.51)$$

that is, the relation (9.44) holds also for $G = E_8$.

Exercise 9.2 Check that the relation (9.44) holds even for $G = E_8 \times E_8$.

We conclude that the Green–Schwarz mechanism can cancel the 10d gauge anomaly for $G = E_8^k$ for all k . But only for $k = 2$ it will cancel *all* anomalies, including the gravitational ones, as we are going to show.

Canceling All Anomalies: Gravitational, Gauge, Mixed

We can generalize the above mechanism to cancel anomalies in terms of a counter-term of the general form

$$\int B^{(2)} X^{(8)}(F, R) \quad (9.52)$$

where $X^{(8)}$ is a $\mathfrak{g} \oplus \mathfrak{so}(10)$ **ad**-invariant polynomial of degree 4 in the curvatures. This counter-term cancels all anomalies whose polynomial has the factorized form $(\text{tr } F^2) X^{(8)}$. In the general case, the $B^{(2)}$ field strength may include, besides the gauge Chern–Simons (CS) term, also a gravitational Chern–Simons term

$$\tilde{H}^{(3)} = dB^{(2)} - c\omega_g^{(3)} - c'\omega_L^{(3)}, \quad (9.53)$$

where c, c' are some constants, and $\omega_g^{(3)}, \omega_L^{(3)}$ are the gauge and Lorentz (i.e. gravitational) Chern–Simons 3-forms

$$\omega_g^{(3)} = \text{tr}\left(AdA + \frac{2}{3}A^3\right), \quad \omega_L^{(3)} = \text{tr}\left(\omega d\omega + \frac{2}{3}\omega^3\right), \quad (9.54)$$

where $\omega \equiv \omega^{(1)}$ is the spacetime spin connection. Under a *local* Lorentz transformation of parameter Θ , the CS form $\omega_L^{(3)}$ changes as

$$\delta\omega_L^{(3)} = d\text{tr}(\Theta R). \quad (9.55)$$

To leave the field strength $\tilde{H}^{(3)}$ invariant, the combined Lorentz and Yang–Mills gauge transformation must be

$$\begin{aligned} \delta A &= D\lambda & \delta\omega &= D\Theta \\ \delta B^{(2)} &= c\text{tr}(\lambda F) + c'\text{tr}(\Theta R). \end{aligned} \quad (9.56)$$

With the gauge transformation (9.56) the counter-term (9.52) cancels an anomaly with polynomial

$$I^{(12)}(F, R) = \left(c\text{tr} F^2 + c'\text{tr} R^2\right)X^{(8)}(F, R). \quad (9.57)$$

It remains to show that the anomaly polynomials in the Type I $SO(32)$ and the two heterotic strings have the factorized form in the RHS of (9.57). The chiral fields are a MW gravitino, a MW dilatino of opposite chirality, and a MW gluino (of the same chirality as the gravitino) in the adjoint of the gauge group. Hence the anomaly polynomial is

$$\begin{aligned} &\frac{1}{2}I_{\text{spin } 3/2}^{(12)}(R) - \frac{1}{2}I_{\text{spin } 1/2}^{(12)}(R) + \frac{1}{2}I_{\text{spin } 1/2}^{(12)}(F, R) = \\ &= \frac{1}{1440} \left\{ -\text{tr}_{\text{ad}}(F^6) + \frac{1}{48}\text{tr}_{\text{ad}}(F^2)\text{tr}_{\text{ad}}(F^4) - \frac{1}{14400}(\text{tr}_{\text{ad}} F^2)^3 \right\} + \\ &\quad + (n - 496) \left(\frac{\text{tr} R^6}{725760} + \frac{\text{tr} R^2 \text{tr} R^4}{552960} + \frac{(\text{tr} R^2)^3}{1327104} \right) + \\ &\quad + \frac{1}{768} Y^{(4)} X^{(8)} \end{aligned} \quad (9.58)$$

where $n = \dim \mathfrak{g}$ and

$$Y^{(4)} = \text{tr } R^2 - \frac{1}{30} \text{tr}_{\text{ad}} F^2 \equiv \text{tr } R^2 - \text{tr } F^2 \quad (9.59)$$

$$X^{(8)} = \text{tr } R^4 + \frac{1}{4} (\text{tr } R^2)^2 - \frac{1}{30} \text{tr}_{\text{ad}} F^2 \text{tr } R^2 + \frac{1}{3} \text{tr}_{\text{ad}} F^4 - \frac{1}{900} (\text{tr}_{\text{ad}} F^2)^2. \quad (9.60)$$

The large brace in Eq. (9.58) vanishes when the gauge group is $SO(32)$ or $E_8 \times E_8$ by the relation (9.44), while both Lie groups have dimension $n = 496$. We remain with a net anomaly polynomial

$$\frac{1}{768} Y^{(4)} X^{(8)} \quad (9.61)$$

which has the factorized form (9.57) with

$$c'/c = -1. \quad (9.62)$$

Note 1 In Sect. 10.1, we shall show by a direct stringy computation that the Lorentz CS term is indeed present in $\tilde{H}^{(3)}$ with $c' = -c$.

Note 2 The gravitational term in $\tilde{H}^{(3)}$ (cf. Eq. (9.53)) was not included in the earlier low-energy effective action (see Sect. 8.6) because it is a higher derivative effect while we truncated the effective Lagrangian to two-derivatives. Indeed the spin connection $\omega^{(1)}$ is proportional to the derivative of the vielbein e^a_μ , so the gravitational CS $\omega_L^{(3)}$ is cubic in derivatives. Its inclusion produces terms in the Lagrangian with more than 2 derivatives which go beyond the usual low-momenta SUGRA truncation.

Topological Constraint The Bianchi identity for the field strength $\tilde{H}^{(3)}$ in (9.53) is

$$d\tilde{H}^{(3)} = c(\text{tr } F^2 - \text{tr } R^2), \quad (9.63)$$

while $\tilde{H}^{(3)}$, being gauge-invariant, is a globally well-defined 3-form. Hence the LHS of (9.63) is cohomologically trivial. This implies that *we may consistently define the string only in backgrounds which satisfy*

$$[\text{tr } F^2] = [\text{tr } R^2]. \quad (9.64)$$

i.e. the Yang–Mills and spin bundles should have equal first Pontragin classes [27]. This equality of Pontragin classes is a kind of “stringy structure” that the background must possess in order for Type I/heterotic string theory to be consistently defined. The connection of this topological restriction on the spacetime configurations to modular invariance is explained in the next section.

9.3 Modular-Invariant $\Rightarrow$ Anomaly-Free

We have checked that all *consistent* 10d superstrings are indeed anomaly-free. From the low-energy effective theory, this looks quite a miracle resting on several magical group theoretical identities; from the higher stringy standpoint these identities are obvious since the theory was quantum consistent to start with. The anomaly is a one-loop effect, and its cancelation should boil down to consistency of one-loop amplitudes, that is, to modular invariance. Besides the five 10d SUSY theories, there are many other string theories which are consistent, e.g. the Narain toroidal compactifications, orbifold constructions, and more generally any world-sheet SCFT with the proper Virasoro central charges. All these models can be seen as the same theory in different “vacua”, and we expect that they are all anomaly-free. However it would be preferable to have a direct and universal proof that modular invariance implies anomaly-freeness which prescind from the details of the world-sheet SCFT. We present such an argument following Schellekens and Warner [31, 32]. The discussion is directly related to Witten’s theory of the *elliptic genus* [33].

For simplicity, we shall consider only the case of a *heterotic* model, that is, the world-sheet theory is supposed to be a $(0, 1)$ SCFT with a d -dimensional Poincaré invariance, with $2 < d \leq 10$, and gauge group G . We may assume d to be *even* since otherwise there are no local anomalies by degree considerations. We write $d = 2n + 2$ with $0 < n \leq 4$. The light-cone world-sheet action contains a decoupled sector with $2n$ free bosons and $2n$ right-moving free fermions (the transverse spacetime coordinates and their superpartners) plus an “internal” $(0,1)$ SCFT theory $\mathcal{T}$ with central charges $(24 - 2n, 12 - 3n)$

$$S = \frac{1}{4\pi} \int d^2z \left(\partial X^i \bar{\partial} X^i + \tilde{\psi}^i \partial \tilde{\psi}_i + \mathcal{L}_{\mathcal{T}} \right) \quad (9.65)$$

where $i = 1, \dots, 2n$ and $\mathcal{L}_{\mathcal{T}}$ is the Lagrangian of the $\mathcal{T}$ theory.

★ Relation to the Elliptic Genus

The partition function of the theory (9.65) takes the general form (same notation as in Chap. 5)

$$\int \frac{d^2\tau}{(\text{Im } \tau)^{n+2}} |\eta(q)|^{-4n} \sum_{\alpha, \beta=0}^1 (Z_{\beta}^{\alpha}(\bar{\tau}))^n T(q, \bar{q})_{\beta}^{\alpha}, \quad (9.66)$$

where $T(q, \bar{q})_{\beta}^{\alpha}$ is the partition function of the theory $\mathcal{T}$ with the appropriate boundary conditions on the right-moving supercurrent $\tilde{T}_F(\bar{z})$ (i.e. spin-structure on the torus). In a consistent string theory, this function is restricted by modular invariance of the integrand in (9.66). Note that

$$\frac{d^2\tau}{(\text{Im } \tau)^{n+2} |\eta(q)|^{4n}} \quad (9.67)$$

is modular-invariant for all $n \in \mathbb{N}$, so the sum

$$\sum_{\alpha, \beta=0}^1 (Z_{\beta}^{\alpha}(\bar{\tau}))^n T(q, \bar{q})_{\beta}^{\alpha} \quad (9.68)$$

should be also modular-invariant in a consistent theory. As shown in detail in Sect. 5.5, the modular group permutes the three even spin-structure partition functions

$$(Z_0^0(\bar{\tau}))^n, \quad (Z_0^1(\bar{\tau}))^n, \quad (Z_1^0(\bar{\tau}))^n, \quad (9.69)$$

up to some phases, so modular invariance requires that the corresponding three functions $T(q, \bar{q})_{\alpha\beta}^{\alpha}$ with $\alpha\beta = 0$ (the even spin-structure ones) are permuted in the same way while picking up the opposite phases. These even spin-structure sectors are of *no* interest for us: they contain only spacetime bosons and non-chiral fermions¹³ and so do not contribute to the anomaly.

The sector $\alpha = \beta = 1$ is the only source of chirality in the theory; the partition function multiplying it in (9.66), $T(q, \bar{q})_1^1 \equiv T(q, \bar{q})$, then contains the gauge and Lorentz representations which contribute to the spacetime anomaly. From **BOX 5.6**, we know that

$$Z_1^1(\bar{\tau} + 1) = e^{-i\pi/6} Z_1^1(\bar{\tau}), \quad Z_1^1(-1/\bar{\tau}) = e^{-i\pi/2} Z_1^1(\bar{\tau}). \quad (9.70)$$

Consistency requires $T(q, \bar{q}) \equiv T(q)$ to depend on q only. Let us review the argument. One has

$$T(q, \bar{q}) = \text{Tr}_R \left[(-1)^{\tilde{F}} q^{L_0 - c/24} \bar{q}^{\tilde{L}_0 - \tilde{c}/24} \right] \quad (9.71)$$

where $\tilde{F}$ is the right-moving Fermi number and the trace is over the right-moving Ramond sector of the (0,1) SCFT $\mathcal{T}$. We claim that only the states with

$$\tilde{L}_0 - \frac{\tilde{c}}{24} \equiv \tilde{G}_0^2 = 0, \quad (9.72)$$

contribute to the trace in (9.71), so that $T(q, \bar{q})$ is a holomorphic function of q (hence of τ). Indeed, if $|\psi\rangle$ is an eigenstate of $\tilde{L}_0 - \tilde{c}/24$ with eigenvalue $\tilde{h} - \tilde{c}/24 \neq 0$, $G_0|\psi\rangle$ is a non-zero eigenvector with the same eigenvalue but opposite $(-1)^{\tilde{F}}$ -parity, and their contributions to the RHS of (9.71) cancel in pairs. The only states contributing to (9.71) are states whose right-movers are in a Ramond vacuum, $\tilde{h} = \tilde{c}/24$, while the left-movers are in an arbitrary state.

The RHS of (9.71) is a special kind of SUSY index called the *elliptic genus* of $\mathcal{T}$ [33]. It can be understood as the index of the Ramond–Dirac operator

$$\sqrt{2} \tilde{G}_0 = P_R^\mu \Gamma_\mu + \text{oscillator contributions} \quad (9.73)$$

computed in the string Hilbert space, i.e. the index of the right-moving Dirac–Ramond operator.

The Grand Anomaly Polynomial We wish to describe the anomaly polynomial $I^{(d+2)}$ in a convenient way which both simplifies the proof of group theoretical identities (that before we checked ‘by hand’) as well as relate the polynomial directly to modular invariance.

Consider the anomaly polynomial generating function

$$A(q, F, R) \stackrel{\text{def}}{=} \hat{A}(R) \sum_{(m,g,l)} q^m \text{Ch}(F, g) \text{Ch}(R, l), \quad (9.74)$$

¹³ Cf. Eq. (5.102): it is obvious that the only term which slips sign under inversion of chirality is the Z_1^1 one which correspond to the *odd* spin-structure.

where the sum is over the mass level m and the gauge and *transverse* Lorentz representations, g and l , which appear in the *left-moving side* of the given heterotic string at mass level m . The right-movers are frozen in the R-vacua and contribute the factor $\hat{A}(R)$. By the relation with the index theorem reviewed in Sect. 9.1, the anomaly polynomial is the $(d+2)$ -form part of the q^0 coefficient of $A(q, F, R)$

$$I(F, R)^{(d+2)} = \left(\text{coeff. of } q^0 \text{ in } A(q, F, R) \right)^{(d+2)}. \quad (9.75)$$

This is obvious for the gauge and mixed anomalies; it is less obvious for the purely gravitational anomalies as far as the contribution of the gravitino and dilatino are concerned: one easily checks that restricting to transverse Lorentz representations and purely transverse Riemann tensors one correctly reproduces these contributions to the anomaly polynomial.¹⁴

The left movers correspond to the spacetime part, $X_L^i(z)$, together with the left movers of $\mathcal{T}$ which are inert under the Lorentz group but carry a current algebra (Kac-Moody Lie algebra) of gauge group G whose charges we denote as J^a . The sum in the RHS of (9.74) factorizes into a contribution from the free oscillators of the fields $X_L^i(z)$ and a contribution from $\mathcal{T}$

$$\sum_{(m,g,l)} q^m \text{Ch}(F, g) \text{Ch}(R, l) = \left(q^{-n/12} \prod_{k=1}^{\infty} \det[1 - q^k e^{iR/2\pi}]^{-1} \right) \times \quad (9.76)$$

$$\times \text{Tr}_{\mathcal{T},R} [(-1)^{F_R} q^{L_0 - (24-2n)/24} e^{iJ \cdot F/2\pi}]$$

Here $\text{Tr}_{\mathcal{T},R}$ means the trace in the right-moving Ramond sector of the (0,1) SCFT $\mathcal{T}$. The factor

$$T(\tau, F) \stackrel{\text{def}}{=} \text{Tr}_{\mathcal{T},R} [(-1)^{F_R} q^{L_0 - (24-2n)/24} e^{iJ \cdot F/2\pi}] \quad (9.77)$$

is the *character valued elliptic genus* of $\mathcal{T}$. Notice that this is the same expression we would get by switching on a flat connection for G (i.e. a 2d Wilson line) along the “time” cycle of the world-sheet torus as we did in Sect. 5.5. For instance, if $\mathcal{T}$ consists of free left-moving fermions, the character valued elliptic genus is

$$\prod_k \eta(\tau)^{-1} \vartheta \left[\begin{matrix} 1/2 \\ 1/2 \end{matrix} \right] \left(\frac{y_k}{(2\pi)^2 i} \mid \tau \right) \equiv T(\tau, F), \quad (9.78)$$

¹⁴ This is a manifestation of the isomorphism of the light-cone and physical covariant spectra.

BOX 9.3 - Modular properties of ϑ -functions with characteristics

From Eq. (2.16) of [34] (paying attention to the different notation) we have

$$\vartheta \left[\begin{smallmatrix} \epsilon/2 \\ \epsilon'/2 \end{smallmatrix} \right] \left(\frac{z}{c\tau + d}, \frac{a\tau + b}{c\tau + d} \right) = \kappa (c\tau + d)^{1/2} \exp \left(\frac{\pi i c z^2}{c\tau + d} \right) \vartheta \left[\begin{smallmatrix} (a\epsilon + c\epsilon' - ac)/2 \\ (b\epsilon + d\epsilon' + bd)/2 \end{smallmatrix} \right] (z, \tau)$$

where κ is a phase given in Eq. (2.17) of [34] and $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$ (so a, c and respectively b, d , are coprime). For the *odd* spin-structure $\epsilon = \epsilon' = 1$, the transformed characteristics in the RHS are again $1/2 \bmod 1$, and using the formula in **BOX 5.5** for the translation of the characteristics by an integer, which simply multiplies the θ -function by a *constant* phase (i.e. a phase which depends on ϵ, ϵ' and a, b, c, d but not on z or τ) and the equation

$$\eta \left(\frac{a\tau + b}{c\tau + d} \right) = (\text{constant phase}) (c\tau + d)^{1/2} \eta(\tau),$$

we get

$$\eta \left(\frac{a\tau + b}{c\tau + d} \right)^{-1} \vartheta \left[\begin{smallmatrix} 1/2 \\ 1/2 \end{smallmatrix} \right] \left(\frac{z}{c\tau + d}, \frac{a\tau + b}{c\tau + d} \right) = (\text{constant phase}) \left(\frac{\pi i c z^2}{c\tau + d} \right) \eta(\tau)^{-1} \vartheta \left[\begin{smallmatrix} 1/2 \\ 1/2 \end{smallmatrix} \right] (z, \tau)$$

where the eigenvalues of F are $\pm i y_k$. Since the elliptic genus is given by a path integral, were not for the 2d gauge anomaly, it would be modular-invariant. The gauge anomaly produces a Gaussian pre-factor

$$T \left(\frac{a\tau + b}{c\tau + d}, \frac{F}{c\tau + d} \right) = (\text{constant phase}) \exp \left(\frac{i k c}{32\pi^3 (c\tau + d)} \text{tr } F^2 \right) T(\tau, F), \quad (9.79)$$

where k is the Kač-Moody level. This formula is easy to check when $\mathcal{T}$ consists of free fermions using the formula in **BOX 9.3**; the general case can be reduced to this one by bosonization of the current algebra.

We compute the first factor in the RHS of Eq. (9.76). The curvature R is in the vector representation of $SO(2n)$; the eigenvalues of $R/2\pi$ have the form $\pm i x_a$ and

$$\begin{aligned} P(q, R) &\equiv q^{-n/12} \prod_{k=1}^{\infty} \det[1 - q^k e^{iR/2\pi}]^{-1} = \\ &= q^{-n/12} \prod_{k=1}^{\infty} \prod_{a=1}^n (1 - q^k e^{x_a})^{-1} (1 - q^k e^{-x_a})^{-1}. \end{aligned} \quad (9.80)$$

The RHS is evaluated by the Jacobi triple product identity (**BOX 5.5**)

$$\prod_{k=1}^{\infty} (1 - q^k) (1 + q^{k-1/2} z) (1 + q^{k-1/2} z^{-1}) = \sum_{n \in \mathbb{Z}} q^{n^2/2} z^n. \quad (9.81)$$

Then

$$\begin{aligned}
 & \prod_{k=1}^{\infty} (1 - q^k e^{x_a})(1 - q^k e^{-x_a}) = \\
 & = (1 - e^{-x_a})^{-1} \prod_{k=1}^{\infty} (1 - q^{k-1/2} (q^{1/2} e^{x_a})(1 - q^{k-1/2} (q^{1/2} e^{x_a})^{-1}) = \\
 & = \frac{q^{-1/8}}{2 \sinh(x_a/2)} \frac{q^{1/24}}{\eta(\tau)} \sum_{n \in \mathbb{Z}} (-1)^n q^{(n+1/2)^2/2} e^{(n+1/2)x_a} \equiv \\
 & \equiv \frac{q^{-1/12}}{2 \sinh(x_a/2)} \frac{1}{\eta(\tau)} \theta_1(x_a/2\pi i | \tau).
 \end{aligned} \tag{9.82}$$

and

$$P(q, R) = \eta(\tau)^n \prod_{a=1}^n \frac{2 \sinh(x_a/2)}{\theta_1(x_a/2\pi i | \tau)} \tag{9.83}$$

$$\widehat{A}(R) P(\tau, R) = \eta(\tau)^n \prod_{a=1}^n \frac{x_a}{\theta_1(x_a/2\pi i | \tau)} \tag{9.84}$$

Hence

$$\begin{aligned}
 & \widehat{A}\left(\frac{R}{c\tau + d}\right) P\left(\frac{a\tau + b}{c\tau + d}, \frac{R}{c\tau + d}\right) = \\
 & = (\text{constant phase}) (c\tau + d)^{-n} \exp\left(\frac{-ic}{32\pi^3(c\tau + d)} \text{tr } R^2\right) \widehat{A}(R) P(\tau, R),
 \end{aligned} \tag{9.85}$$

and referring back to Eq. (9.76)

$$\begin{aligned}
 & A\left(\frac{a\tau + b}{c\tau + d}, \frac{R}{c\tau + d}, \frac{F}{c\tau + d}\right) = \\
 & = (\text{phase}) (c\tau + d)^{-n} \exp\left(\frac{ic}{32\pi^3(c\tau + d)} (k \text{tr } F^2 - \text{tr } R^2)\right) A(\tau, R, F).
 \end{aligned} \tag{9.86}$$

Specializing this equation at $F = R = 0$, we get the condition of modular invariance of the 1-loop partition function in the trivial background. Conversely, modular invariance implies this equation: the only difference is that now we have an exponential pre-factor which is quadratic in the field strengths. The argument also implies that in a modular-invariant theory *the constant phase should be trivial*.

Modular invariance is preserved in a non-trivial background iff the exponential pre-factor is trivial, i.e. 1 in cohomology

$$k[\text{tr } F^2] - [\text{tr } R^2] = 0, \quad (9.87)$$

which is the same consistency condition we got before for the 10d heterotic strings (in that case $k = 1$).

It remains to show that the modular invariance condition (9.86) implies absence of anomalies under the above consistency condition.

Theorem 9.1 *Assume:*

1. *the modular invariance condition (9.86) holds with trivial constant phase;*
2. *the background consistency condition (9.87).*

Then the anomaly polynomial vanishes

$$I^{(d+2)} \equiv \left(\text{coeff. of } q^0 \text{ in } A(q, R, F) \right)^{(2(n+2))} = 0. \quad (9.88)$$

The proof is based on the following

Lemma 9.1 *Let $f(\tau)$ be a function which is holomorphic in the upper half-plane $\mathcal{H}$, possibly with poles at infinity, and modular of weight 2, i.e. [37–39]*

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^2 f(\tau) \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}). \quad (9.89)$$

Let

$$f(\tau) = \sum_{k \geq -h} q^k f_k \quad (9.90)$$

($q \equiv e^{2\pi i \tau}$) be its Fourier expansion. Then the coefficient of q^0 vanishes, $f_0 = 0$.

Proof Let $j(\tau)$ be the standard modular-invariant function defined in **BOX 9.4**. Under modular transformation the derivative changes as

$$\frac{d}{d\tau} \rightarrow (c\tau + d)^2 \frac{d}{d\tau}. \quad (9.91)$$

Then, for all polynomials $P(\cdot)$, the function

$$\frac{d}{d\tau} P(j), \quad (9.92)$$

is holomorphic in $\mathcal{H}$ and modular of weight 2. We can find a polynomial $P(\cdot)$ such that the weight 2 holomorphic modular function

$$f(\tau) - \frac{d}{d\tau} P(j) \quad (9.93)$$

has no poles at infinity and hence vanishes because there is no non-zero weight 2 modular function which is holomorphic everywhere including $i\infty$ [39]. Let $P(j) = \sum_k q^k P_k$ be the Fourier series of the modular-invariant polynomial. We have

$$f_k = 2\pi i k P_k \quad \text{and, in particular, } f_0 = 0. \quad (9.94)$$

Proof (of Theorem) Consider the $(d+2)$ form component of $A(\tau, R, F)$. It has the general structure

BOX 9.4 - The modular function $j(\tau)$ as a partition function

The moduli space of elliptic curves, $\mathcal{H}/SL(2, \mathbb{Z})$, has a natural compactification $\overline{\mathcal{H}/SL(2, \mathbb{Z})}$ obtained by adding the point at infinity $i\infty$. $\overline{\mathcal{H}/SL(2, \mathbb{Z})}$ is a compact complex manifold of dimension 1: being simply-connected^a is isomorphic to the Riemann sphere. The isomorphism $\overline{\mathcal{H}/SL(2, \mathbb{Z})} \rightarrow \mathbb{P}^1$ is given by a meromorphic function $f: \mathcal{H} \rightarrow \mathbb{P}^1$ which is modular invariant

$$f\left(\frac{a\tau + b}{c\tau + d}\right) = f(\tau) \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$$

of degree 1, i.e. with a single pole of order 1 in $\overline{\mathcal{H}/SL(2, \mathbb{Z})}$. All degree-1 modular-invariant functions are obtained by any one of them, $j(\tau)$, by $SL(2, \mathbb{C})$ automorphisms of the sphere

$$f(\tau) = \frac{aj(\tau) + b}{cj(\tau) + d}, \quad \text{for } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{C}) \quad (\spadesuit)$$

The standard choice is the modular-invariant function $j(\tau)$ which is holomorphic in the fundamental domain $\mathcal{H}/SL(2, \mathbb{Z})$ with a simple pole at $\tau = i\infty$ (see, e.g. [36–38])

$$j(\tau) = \frac{1}{q} + 744 + 196884q + \dots \quad \text{where } q = e^{2\pi i\tau} \quad (\dagger)$$

$$j\left(\frac{a\tau + b}{c\tau + d}\right) = j(\tau) \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}) \quad (\ddagger)$$

In view of $(\spadesuit)$ the first two terms in the expansion $(\dagger)$ of j around $q = 0$ identify uniquely the modular-invariant function $j(\tau)$. The function $j(\tau)$ can be written in many ways; the most illuminating one is in terms of free fermion partition functions $Z^\alpha_\beta(\tau)$ defined in Eq. (5.71)

$$j(\tau) = \left(\frac{1}{2} \left(Z^0_0(\tau)^8 + Z^1_0(\tau)^8 + Z^0_1(\tau)^8 + Z^1_1(\tau)^8 \right) \right)^3$$

which is the partition function of 48 free left-moving MW fermions, divided in three blocks of 16, with an independent GSO projection in each block, i.e. the partition function of 24 free compact scalars whose momenta take value in the even self-dual lattice $\Gamma_8 \oplus \Gamma_8 \oplus \Gamma_8$. In other words, $j(\tau)$ is the partition function of the purely left-moving $E_8 \times E_8 \times E_8$ Kač-Moody current algebra at level 1. From the modular transformations of the lattice partition functions computed in Chaps. 5 and 7, we see that $j(\tau)$ is modular-invariant. Its expansion around $q = 0$ is

$$q^{-c/24} \left(1 + \#(h = 1 \text{ currents})q + O(q^2) \right) = \frac{1}{q} + 3 \dim E_8 + O(q) \equiv \frac{1}{q} + 774 + O(q)$$

which uniquely identifies the modular-invariant partition function with $j(\tau)$.

In **Fact 5.1**, we saw that the partition function of all purely left-moving CFT with $(c, \tilde{c}) = (24n, 0)$ is modular-invariant iff its operator algebra $\mathfrak{A}$ is maximal local with integral spins. The CFT whose partition function is $j(\tau)$ (more generally $j(\tau)^n$) is an example of this situation; we have $(c, \tilde{c}) = (24, 0)$ while the GSO projection yields maximal locality of $\mathfrak{A}$ and integral spins

^a $SL(2, \mathbb{Z})$ is generated by two elements S and ST of finite order. By the Cartan–Hadamard theorem an element of finite order in $\Gamma \subset SL(2, \mathbb{R})$ has a fixed point in $\mathcal{H}$ hence it maps to the identity in $\pi_1(\mathcal{H}/\Gamma)$. For $\Gamma = SL(2, \mathbb{Z})$ this gives $\pi_1 = 1$. [35],

$$A(\tau, R, F)^{(2(n+2))} = \sum_{\ell} f(\tau)_{\ell} X_{\ell}^{(2(n+2))}, \quad (9.95)$$

where $X_{\ell}^{(2(n+2))}$ are invariant polynomials in the curvatures R, F of degree $n+2$. Under the assumption (9.87), Eq. (9.86) implies

$$f\left(\frac{a\tau+b}{c\tau+d}\right)_{\ell} = (c\tau+d)^2 f(\tau)_{\ell}. \quad (9.96)$$

By the **Lemma** the coefficient of q^0 in each function $f(q)_{\ell}$ vanishes; the **Theorem** follows. $\square$

References

1. 't Hooft, G.: Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. NATO Sci. Ser. B. **59**, 135–157 (1980)
2. Zumino, B.: Chiral anomalies and differential geometry. In: DeWitt, B.S., Stora, R. (eds.) *Relativité, groupes et topologie II*, Les Houches, session XL, 27 juin–4 août 1983. North-Holland (1984)
3. Alvarez-Gaume, L., Witten, E.: Gravitational anomalies. Nucl. Phys. B. **234**, 269 (1984)
4. Alvarez, O., Singer, I.M., Zumino, B.: Gravitational anomalies and the family's index theorem. Commun. Math. Phys. **96**, 409 (1984)
5. Alvarez-Gaume, L., Ginsparg, P.H.: The topological meaning of nonabelian anomalies. Nucl. Phys. B. **243**, 449–474 (1984)
6. Zumino, B., Wu, Y.S., Zee, A.: Chiral anomalies, higher dimensions, and differential geometry. Nucl. Phys. B. **239**, 477–507 (1984)
7. Bardeen, W.A., Zumino, B.: Consistent and covariant anomalies in gauge and gravitational theories. Nucl. Phys. B. **244**, 421–453 (1984)
8. Alvarez-Gaumé, L., Ginsparg, P.H.: The structure of gauge and gravitational anomalies. Ann. Phys. **161**, 423 (1985) [erratum: Ann. Phys. **171**, 233 (1986)]
9. Manes, J., Stora, R., Zumino, B.: Algebraic study of chiral anomalies. Commun. Math. Phys. **102**, 157 (1985)
10. Bertlmann, R.A.: *Anomalies in Quantum Field Theory*. Oxford University Press (2000)
11. Wess, J., Zumino, B.: Consequences of anomalous Ward identities. Phys. Lett. B. **37**, 95 (1971)
12. Bourbaki, N.: *Groupes et algèbres de Lie, Chapitres 7 et 8, Éléments de Mathématique*. Springer (1981)
13. Knapp, A.W.: *Lie Groups Beyond an Introduction*. Birkhäuser (1996)
14. Husemoller, D.: *Fibre Bundles*. Graduate Texts in Mathematics, vol. 20, 3rd edn. Springer (1994)
15. Milnor, J.W., Stasheff, J.D.: *Characteristic Classes*. Princeton University Press (1974)
16. Bott, R., Tu, L.W.: *Differential Forms in Algebraic Topology*. Graduate Texts in Mathematics, vol. 82. Springer (1982)
17. Chern, S.S.: *Complex Manifolds without Potential Theory* (with an Appendix in the Geometry of Characteristic Classes). Springer (1979)
18. Atiyah, M.F., Singer, I.M.: The index of elliptic operators in compact manifolds. Bull. Am. Math. Soc. **69**, 422–433 (1963)
19. Atiyah, M.F., Singer, I.M.: The index of elliptic operators, I. Ann. Math. **87**, 484–530 (1968)
20. Atiyah, M.F., Singer, I.M.: The index of elliptic operators, II. Ann. Math. **87**, 531–545 (1968)
21. Atiyah, M.F., Singer, I.M.: The index of elliptic operators, III. Ann. Math. **87**, 546–604 (1968)
22. Atiyah, M.F., Singer, I.M.: The index of elliptic operators, IV. Ann. Math. **93**, 119–138 (1971)
23. Atiyah, M.F., Singer, I.M.: Dirac operators coupled to vector potentials. Proc. Nat. Acad. Sci. USA **81**, 2597–2600 (1984)
24. Alvarez-Gaumé, L.: Supersymmetry and the Atiyah-Singer index theorem. Commun. Math. Phys. **90**, 161–173 (1983)

25. Berline, N., Getzler, E., Vergne, M.: Heat kernels and Dirac Operators. Springer (2004)
26. Romer, H.: Axial anomaly and boundary terms for general spinor fields. Phys. Lett. **82B**, 172 (1979)
27. Hirzebruch, F.: Topological methods in Algebraic Geometry. Classical in Mathematics. Springer (1995)
28. Green, M.B., Schwarz, J.H.: Anomaly cancellation in supersymmetric $D = 10$ gauge theory and superstring theory. Phys. Lett. B. **149**, 117–122 (1984)
29. Atiyah, M.F., Tall, D.O.: Group representations, λ -rings and the J -homomorphism. Topol. **8**, 253–297 (1969)
30. Fuchs, J., Schweigert, C.: Symmetries, Lie Algebras and Representations. A graduate course for physicists. Cambridge Monographs on Mathematical Physics. Cambridge University Press (1997)
31. Schellekens, A.N., Warner, N.P.: Anomalies, characters, and strings. Nucl. Phys. B. **287**, 317–361 (1987)
32. Schellekens, A.N., Warner, N.P.: Anomalies and modular invariance in string theory. Phys. Lett. B. **177**, 317–323 (1986)
33. Witten, E.: Elliptic genera and quantum field theory. Commun. Math. Phys. **109**, 525–536 (1987)
34. Farkas, H.M., Kra, I.: Theta constants, Riemann Surfaces and the Modular Group. Graduate Studies in Mathematics, vol. 37. AMS (2001)
35. Postnikov, M.M.: Geometry VI. Riemannian Geometry. Encyclopaedia Mathematica Sciences, vol. 91. Springer (2001)
36. NIST Digital Library of Mathematical Functions. <https://dlmf.nist.gov>
37. Bruinier, J.H., van der Geer, G., Harder, G., Zagier, D.: The 1-2-3 of Modular Forms, Lectures at a Summer School in Nordfjordeid, Norway. Universitext. Springer (2008)
38. Diamond, F., Shurman, J.: A First Course in Modular Forms. Graduate Texts in Mathematics, vol. 228. Springer (2005)
39. Serre, J.-P.: A Course in Arithmetics. Springer (1973)

Chapter 10

Superstring Amplitudes

Non-Renormalization Theorems



Abstract In this chapter we discuss general aspects of g -loop amplitudes in superstring theories, outlining the intrinsic geometrical approach in terms of super-Riemann surfaces, and describing computational tools. We compute in detail some sample superstring amplitudes with four aims in mind: (a) to illustrate issues and techniques, (b) to introduce some useful tricks, (c) to check that the CS couplings and counter-terms required by the Green–Schwarz anomaly cancelation mechanism are actually present in the physical heterotic superstrings, and (d) to give an explicit perturbative proof (first at one-loop then to all-loops) of the SUSY non-renormalization theorems we inferred in Chap. 8 on general grounds.

This chapter is technical. The reader may prefer to skip it in a first reading.

10.1 Tree-Level Amplitudes

Tree-level amplitudes are easy to compute. All we need are the expectation values on the sphere¹ of the product of vertices with the appropriate number of bosonic/fermionic coordinates fixed to get rid of the residual superconformal symmetry related to the zero-modes of the ghosts c , $\tilde{c}$, γ , and $\tilde{\gamma}$ (cf. Chap. 1).

Type I Disk Amplitudes: Three Massless Vectors

The Type I disk amplitude for three open-sector massless bosons is²

$$\frac{1}{\alpha' g_0^2} \left\langle c V_1^{(-1)}(x_1) c V_2^{(-1)}(x_2) c V_3^{(0)}(x_3) \right\rangle_{\mathcal{H}} + (V_1 \leftrightarrow V_2) \quad (10.1)$$

where the disk is identified with the upper half-plane $\mathcal{H}$ whose boundary is the real axis plus the point $i\infty$ and we take $x_1 > x_2 > x_3$. The total picture of the vertices is -2 to absorb the two real zero-modes of the bosonic ghosts. The overall normalization $1/\alpha' g_0^2$ is as in corresponding bosonic string amplitude.

¹ For Type I also on the disk and $\mathbb{RP}^2$.

² As always a superscript in parenthesis on a vertex, $V^{(q)}$ means that the vertex is in picture q .

Exercise 10.1 Show that the proper normalization of the massless NS vertices is

$$V^{(-1)} = g_0 t^a e_\mu \psi^\mu e^{-\phi} e^{ikX}, \quad V^{(0)} = g_0 (2/\alpha')^{-1/2} t^a e_\mu (\psi^\mu u + 2\alpha' k_\nu \psi^\nu \psi^\mu) e^{ikX}.$$

The relevant disk correlation functions are ($x_{ij} \equiv x_i - x_j$)

$$\langle \psi^\mu(x_1) \psi^\nu(x_2) \rangle = \frac{\eta^{\mu\nu}}{x_{12}}, \quad \langle e^{-\phi(x_1)} e^{-\phi(x_2)} \rangle = \frac{1}{x_{12}} \quad (10.2)$$

$$\langle c(x_1) c(x_2) c(x_3) \rangle = x_{12} x_{13} x_{23}, \quad (10.3)$$

$$\begin{aligned} & \left\langle \psi^\mu e^{ik_1 \cdot X}(x_1) \psi^\nu e^{ik_2 \cdot X}(x_2) (i \dot{X}^\rho + 2\alpha' k_3 \cdot \psi \psi^\rho) e^{ik_3 \cdot X}(x_3) \right\rangle = \\ & = 2i\alpha' (2\pi)^{10} \delta\left(\sum_i k_i\right) \left(-\frac{\eta^{\mu\nu} k_1^\rho}{x_{12} x_{13}} - \frac{\eta^{\mu\nu} k_2^\rho}{x_{12} x_{23}} + \frac{\eta^{\mu\rho} k_3^\nu - \eta^{\nu\rho} k_3^\mu}{x_{13} x_{23}} \right). \end{aligned} \quad (10.4)$$

In addition, the amplitude is multiplied by the Chan–Paton factor

$$\text{tr}(t^{a_1} t^{a_2} t^{a_3}) \equiv \frac{1}{2} \text{tr}([t^{a_1}, t^{a_2}] t^{a_3}), \quad (10.5)$$

where the trace is in the $SO(32)$ vector representation and we used that $SO(32)$ has no symmetric **ad**-invariant of degrees 1 or 3 so $\text{tr}(\{t^{a_1}, t^{a_2}\} t^{a_3}) = 0$. Summing the two terms in Eq. (10.1) we get the amplitude

$$i g_{\text{YM}} (2\pi)^{10} \delta\left(\sum_i k_i\right) \text{tr}([t^{a_1}, t^{a_2}] t^{a_3}) \left(\eta^{\mu\nu} (k_1 - k_2)^\rho - 2\eta^{\nu\rho} k_3^\mu + 2\eta^{\rho\mu} k_3^\nu \right) \quad (10.6)$$

where, keeping track of the vertex normalization, $g_{\text{YM}} \equiv g_0 (2\alpha')^{-1/2}$. Since the amplitude is contracted with transverse polarizations, $k_i^\mu e_{i\mu} = 0$, we are free to add terms proportional to k_1^μ or k_2^ν . We add inside the big parenthesis the expression

$$- \eta^{\nu\rho} k_1^\mu + \eta^{\rho\mu} k_2^\nu \equiv \eta^{\nu\rho} (k_2 + k_3)^\mu - \eta^{\rho\mu} (k_3 + k_1)^\nu, \quad (10.7)$$

to get the amplitude in its symmetric form

$$i g_{\text{YM}} (2\pi)^{10} \delta\left(\sum_i k_i\right) \text{tr}([t^{a_1}, t^{a_2}] t^{a_3}) V^{\mu\nu\rho} e_{1\mu} e_{2\nu} e_{3\rho} \quad (10.8)$$

$$V^{\mu\nu\rho} \stackrel{\text{def}}{=} \eta^{\mu\nu} (k_1 - k_2)^\rho + \eta^{\nu\rho} (k_2 - k_3)^\mu + \eta^{\rho\mu} (k_3 - k_1)^\nu. \quad (10.9)$$

This is the ordinary Yang–Mills 3-vertex.³ Unlike the bosonic amplitude, Eq. (4.138), there is no k^3 term, hence no F^3 term in the low-energy effective action. It is known that such a term is not consistent with 10d $\mathcal{N} = 1$ SUSY. Then the F^3 term cannot be generated by higher-loop corrections: this is an instance of a non-renormalization theorem that we shall prove from a stringy perspective in Sect. 10.4.

³ The amplitude vanishes on-shell for real momenta, but its complex continuation is non-trivial.

Type I Disk Amplitudes: Two Fermions, One Boson

The relevant CFT amplitudes now are

$$\langle e^{-\phi(x_1)/2} e^{-\phi(x_2)/2} e^{-\phi(x_3)} \rangle = \frac{1}{x_{12}^{1/4} x_{13}^{1/2} x_{23}^{1/2}} \quad (10.10)$$

$$\langle S_\alpha(x_1) S_\beta(x_2) \rangle = \frac{C_{\alpha\beta}}{x_{12}^{5/4}} \quad (10.11)$$

$$\langle S_\alpha(x_1) S_\beta(x_2) \psi^\mu(x_3) \rangle = \frac{1}{\sqrt{2} x_{12}^{3/4} x_{13}^{1/2} x_{23}^{1/2}} \Gamma_{\alpha\beta}^\mu, \quad (10.12)$$

as we easily see using the $SO(10)$ current algebra. In 10d the three-point function (10.12) is non-zero only for two spinors of the same chirality (we raise/lower spinor indices with the charge conjugation matrix $C_{\alpha\beta}$). The Majorana condition yields

$$\Gamma_{\alpha\beta}^\mu = -\Gamma_{\beta\alpha}^\mu. \quad (10.13)$$

The amplitude is then proportional to

$$\text{tr}(t^{a_1} t^{a_2} t^{a_3}) \Gamma_{\alpha\beta} + \text{tr}(t^{a_2} t^{a_1} t^{a_3}) \Gamma_{\beta\alpha} \equiv \text{tr}([t^{a_1}, t^{a_2}] t^{a_3}) \Gamma_{\alpha\beta}^\mu. \quad (10.14)$$

Exercise 10.2 Show that the proper normalization of the gaugino vertex, in the $-\frac{1}{2}$ picture, is $g_0(\alpha')^{1/4} e^{-\phi/2} S_\alpha e^{ik \cdot X}$.

Inserting all factors and normalizations, the amplitude becomes

$$i g_{\text{YM}} (2\pi)^{10} \delta\left(\sum_i k_i\right) \text{tr}([t^{a_1}, t^{a_2}] t^{a_3}) \Gamma_{\alpha\beta}^\mu, \quad (10.15)$$

which is the two gluini, one gluon tree-level amplitude in 10d SYM.

Heterotic Sphere Amplitudes

The closed string three-point amplitudes on the sphere are products of left/right amplitudes. For the heterotic string we need the two- and three-current functions

$$\langle J^a(z_1) J^b(z_2) \rangle = \frac{\hat{k} \delta^{ab}}{z_{12}^2} \quad (10.16)$$

$$\langle J^a(z_1) J^b(z_2) J^c(z_3) \rangle = \frac{i \hat{k} f^{abc}}{z_{12} z_{13} z_{23}}. \quad (10.17)$$

To normalize the two-point function we multiply each vertex operator by $\hat{k}^{-1/2}$. For the 10d heterotic string $2\hat{k}/\psi^2 \equiv k = 1$. We normalize the product in the gauge algebra $\mathfrak{g}$ to be $1/h$ times the trace in the adjoint; this gives $\psi^2 = 1$ i.e. $\hat{k} = \frac{1}{2}$. Hence the normalization of the current algebra three-point functions is

$$i \hat{k}^{-1/2} f^{abc} = \sqrt{2} \text{tr}([t^a, t^b] t^c). \quad (10.18)$$

The same result can be obtained by using the free fermionic form of the current, $2^{-1/2} i t_{AB}^a \lambda^A \lambda^B$ or the free bosonic form. Another useful CFT expectation value is

$$\left\langle \prod_{i=1}^3 i e_i \cdot \partial X e^{i k_i \cdot X}(z_i, \bar{z}_i) \right\rangle = \frac{(\alpha')^2}{8} \frac{1}{z_{12} z_{13} z_{23}} T^{\mu\nu\rho} e_{1\mu} e_{2\nu} e_{3\rho} \quad (10.19)$$

where

$$T^{\mu\nu\rho} = k_{23}^\mu \eta^{\nu\rho} + k_{31}^\nu \eta^{\rho\mu} + k_{12}^\rho \eta^{\mu\nu} + \frac{\alpha'}{2} k_{23}^\mu k_{31}^\nu k_{12}^\rho, \quad (10.20)$$

is the amplitude in the bosonic string, Eq. (4.139). To write the expression in this form we used the mass-shell condition $k_i^2 = 0$ and transversality $e_i \cdot k_i = 0$.

Three-Gauge Bosons The left-moving part of the heterotic 3-gauge boson amplitude is the 3-point function of the currents (times the 3 c -ghost function), Eq. (10.18), while the right-moving part is like the open superstring amplitude computed in Eq. (10.8). Including the overall factor $8\pi/\alpha' g_c^2$ (as in the bosonic string) the heterotic three-gauge bosons amplitude becomes

$$4\pi g_c (\alpha')^{-1/2} i (2\pi)^{10} \delta\left(\sum_i k_i\right) e_{1\mu} e_{2\nu} e_{3\rho} V^{\mu\nu\rho} \text{tr}([t^a, t^b] t^c). \quad (10.21)$$

Up to the definition of the YM coupling, this is the same as the open superstring amplitude, and there is no k^3 term in agreement with 10d $\mathcal{N} = 1$ supersymmetry.

Three Massless Neutral Bosons (Graviton, Dilaton, 2-Form) The amplitude is

$$\pi i g_c (2\pi)^{10} \delta\left(\sum_i k_i\right) e_{1\mu\sigma} e_{2\nu\omega} e_{3\rho\lambda} V^{\mu\nu\rho} T^{\sigma\omega\lambda}, \quad (10.22)$$

with the tensors $V^{\mu\nu\rho}$, $T^{\mu\nu\rho}$ defined in Eqs. (10.9), (10.20). One can relate the coupling g_c to the YM coupling and the gravitational coupling κ ; the relation is (of course) the same one obtained by the analysis of the low-energy theory in Chap. 8.

Specializing the amplitude (10.22) to two symmetric and one antisymmetric polarization, we see that there is an order k^4 interaction of two gravitons and one B -field. An effective action coupling built out of curvatures and field strengths has at least 5 derivatives. The interaction must be the gravitational Chern–Simons coupling

$$(dB)^{(3)} \wedge *\omega_L^{(3)} \quad (10.23)$$

which in the heterotic string was required to cancel the anomalies, cf. Sect. 9.2. This shows that the Lorentz CS coupling implied by the Green-Schwarz mechanism is actually present in string theory.

Two Gauge, One Neutral Massless Boson The amplitude is

$$\pi i g_c (2\pi)^{10} \delta\left(\sum_i k_i\right) e_{1\mu\sigma} e_{2\nu} e_{3\rho} k_{23}^\sigma V^{\mu\nu\rho} \delta^{ab}. \quad (10.24)$$

Indeed, the right-moving part is as before and the left-moving one is proportional to

$$\langle c(z_1)c(z_2)c(z_3) \rangle \langle J^a(z_2)J^b(z_3) \rangle \langle \partial X^\sigma e^{ik_1 \cdot X}(z_1) e^{ik_2 \cdot X}(z_2) e^{ik_3 \cdot X}(z_3) \rangle, \quad (10.25)$$

while (up to overall normalization and momentum δ -function)

$$\langle \partial X^\sigma e^{ik_1 \cdot X}(z_1) e^{ik_2 \cdot X}(z_2) e^{ik_3 \cdot X}(z_3) \rangle \sim \frac{ik_2^\sigma}{z_{12}} + \frac{ik_3^\sigma}{z_{13}}. \quad (10.26)$$

We are free to add terms proportional to $k_1^\sigma \equiv -(k_2 + k_3)^\sigma$ since $e_{1\mu\sigma}k_1^\sigma = 0$

$$\frac{ik_2^\sigma}{z_{12}} + \frac{ik_3^\sigma}{z_{13}} - \frac{i}{2} \left(\frac{1}{z_{12}} + \frac{1}{z_{13}} \right) (k_2 + k_3)^\sigma = \frac{i}{2} \frac{z_{23}}{z_{12}z_{13}} (k_2 - k_3)^\sigma. \quad (10.27)$$

All factors of z_{ij} cancel, and we remain with (10.24).

Consider the part of the amplitude in Eq. (10.24) for the 2-form field, i.e.

$$\pi i g_c (2\pi)^{10} \delta \left(\sum_i k_i \right) e_{1[\mu\sigma]} e_{2\nu} e_{3\rho} k_{23}^{[\sigma} V^{\mu]\nu\rho} \delta^{ab}, \quad (10.28)$$

which is proportional to

$$e_{1[\mu\sigma]} k_{1\rho} e_2^{[\rho} k_3^\sigma e_3^{\mu]}, \quad (10.29)$$

and corresponds to an interaction of the form $*dB \wedge \omega_{\text{YM}}^{(3)}$, where now $\omega_{\text{YM}}^{(3)}$ is the YM Chern-Simons 3-form. Thus we recover that the field strength $\tilde{H}$ is improved by a term proportional to the gauge Chern-Simons term $\omega_{\text{YM}}^{(3)}$ as required by spacetime supersymmetry and the Green–Schwarz mechanism.

Type I/II Sphere Amplitudes

In Type I/Type II theory the amplitude of 3 massless NS-NS bosons on $\Sigma \equiv S^2$ is

$$\pi i g_c (2\pi)^{10} \delta \left(\sum_i k_i \right) e_{1\mu\sigma} e_{2\nu\omega} e_{3\rho\lambda} V^{\mu\nu\rho} V^{\sigma\omega\lambda}, \quad (10.30)$$

where $V^{\mu\nu\rho}$ is as in Eq. (10.9). The normalization $8\pi/g_c^2\alpha'$ and the relation $\kappa = 2\pi g_c$ are the same as in closed bosonic string.

Remark 1 The tensor structure in Type I/II, Eq. (10.30), is simpler than the heterotic one, Eq. (10.22), since there are only terms of order k^2 but not of order k^4 (nor k^6 as in the bosonic string (4.158)). Thus there is no (Riemann)² or (Riemann)³ corrections in Type I/II and no (Riemann)³ in the heterotic case cf. Eq. (10.22). In Type II there is no Lorentz CS either. This was expected: Type II has $\mathcal{N} = 2$ space-time SUSY which forbids more couplings than the weaker $\mathcal{N} = 1$ SUSY.⁴ This observation implies a SUSY non-renormalization theorem which remains valid to all orders.

⁴ In Type I we have an “accidental” cancelation of the tree-level (Riemann)² correction which is not a consequence of SUSY.

Remark 2 The previous remark shows that—contrary to the heterotic case—the field strength of $B_{\mu\nu}$ is not improved in Type II. Indeed no CS term is required by anomaly cancelation in this case. In Type I the 2-form whose field strength gets improved to cancel the anomaly comes from the R-R sector not the NS-NS one.

10.2 General Amplitudes

The computation of higher-loop contributions in superstring theory is very subtle. We had a hint of this fact when discussing the β , γ system, its bosonization, and picture-changing in Chaps. 2 and 3. We shall be rather sketchy; for details see the original literature [1–6] or Witten’s modern revisitation [7–11].

There are two main approaches. One is very geometric and intrinsic: it is based on the theory of *super-Riemann surfaces* (SRS). The second one is more physics-style and slightly ad hoc: it is based on the notion of absorbing the ghosts’ zero-modes and the related insertion of *picture changing operators* (PCO).

Super-Riemann Surfaces

The bosonic string higher-loop amplitudes were sketched in Chap. 1. In the bosonic case a g -loop amplitude with n operator insertions is given by an integral over the moduli space $\mathcal{M}_{g,n}$ of conformal structures on a genus- g surface with n -punctures. Putting a conformal structure on an orientable smooth surface Σ is equivalent to prescribing the equivalence class of its 2d metric g_{ab} modulo the $\mathbf{Diff}^+ \times \mathbf{Weyl}$ gauge transformations which fix the punctures. We need the corresponding statements in the SUSY case. For simplicity we consider the vacuum amplitude ($n = 0$) and focus on $g \geq 2$.

Let Σ a compact, orientable 2d manifold of genus $g \geq 2$. A *superconformal* structure on Σ is an equivalence class of metric and gravitino fields, (g_{ab}, χ_a) , modulo $\mathbf{super-Diff}^+ \times \mathbf{super-Weyl}$; this entails a choice of spin-structure $\mathcal{L}$. The superconformal structure is precisely the gauge-invariant content of the 2d field configuration (g_{ab}, χ_a) , and the physical amplitude is an integral over all superconformal structures parametrized by the superconformal moduli space $\mathcal{SM}_g$. $\mathcal{SM}_g$ is a complex supermanifold⁵ of dimension $3(g-1)|2(g-1)$: indeed we know from Chap. 1 that the tangent superspace $T\mathcal{SM}_g$ is

$$(T\mathcal{SM}_g)_{\text{even}} \simeq H^1(\Sigma, K^{-1}), \quad (T\mathcal{SM}_g)_{\text{odd}} \simeq H^1(\Sigma, \mathcal{L}^{-1}). \quad (10.31)$$

Mathematically, a superconformal structure on a complex $1|1$ supermanifold S is the datum of a sub-super-bundle $\mathfrak{D} \subset TS$, of complex rank $0|1$, which is *maximally non-integrable*: if $v \in \mathfrak{D}$ is a section which is nowhere zero in an open set $U \subset S$, then v^2 is a section of $TS|_U$ with $v^2 \not\equiv 0 \bmod \mathfrak{D}$ pointwise [7]. Concretely, a superconformal

⁵ A complex supermanifold of dimension $m|n$ is a superspace locally modeled on $\mathbb{C}^{m|n}$, the superspace with m even ($\equiv$ bosonic) coordinates $x_i \in \mathbb{C}^m$ and n odd ($\equiv$ fermionic) complex coordinates θ_a . For background on supermanifolds see [9, 12–15].

structure is an isomorphism class of super-atlases for S whose transition functions satisfy Eq. (10.34) below. One covers S by coordinate patches U_α with even z_α and odd θ_α local complex coordinates, and defines the superderivative

$$D_{\theta_\alpha} = \frac{\partial}{\partial \theta_\alpha} + \theta_\alpha \frac{\partial}{\partial z_\alpha}, \quad \Rightarrow \quad D_{\theta_\alpha}^2 = \partial_{z_\alpha}. \quad (10.32)$$

In the intersection $U_\alpha \cap U_\beta$ of two coordinate patches

$$z_\alpha = f_{\alpha\beta}(z_\beta, \theta_\beta), \quad \theta_\alpha = \phi_{\alpha\beta}(z_\beta, \theta_\beta), \quad (10.33)$$

where $f_{\alpha\beta}$ (resp. $\phi_{\alpha\beta}$) is an analytic even (resp. odd) function of its arguments. Such a super-atlas defines a superconformal structure iff the D_{θ_α} 's generate a global super-bundle $\mathfrak{D} \subset TS$; the second Eq. (10.32) says that $\mathfrak{D}$ is maximally non-integrable. The super-bundle $\mathfrak{D}$ is well-defined iff D_{θ_α} is proportional to D_{θ_β} in $U_\alpha \cap U_\beta$ i.e. iff

$$D_{\theta_\beta} z_\alpha = \theta_\alpha D_{\theta_\beta} \theta_\alpha \quad \Rightarrow \quad D_{\theta_\beta} = (D_{\theta_\beta} \theta_\alpha) D_{\theta_\alpha}, \quad (10.34)$$

whose general solution is [16]

$$z_\alpha = h_{\alpha\beta}(z_\beta) + \theta_\beta (\partial_{z_\beta} h_{\alpha\beta}) \epsilon_{\alpha\beta}(z_\beta), \quad (10.35)$$

$$\theta_\alpha = (\partial_{z_1} h_{\alpha\beta})^{1/2} (\epsilon_{\alpha\beta}(z_\beta) + \theta_\beta + \frac{1}{2} \theta_\beta \epsilon_{\alpha\beta} \partial_{z_\beta} \epsilon_{\alpha\beta}(z_\beta)), \quad (10.36)$$

where $h_{\alpha\beta}$ ($\epsilon_{\alpha\beta}$) is an even (odd) analytic function of z_β . A complex 1|1 supermanifold S with a superconformal structure is called a *super-Riemann surface* (SRS). If we forget the θ_α 's in the definition of S , we get an ordinary Riemann surface, with local coordinates z_α and transition functions $h_{\alpha\beta}(z_\beta)$, together with a choice of square root of $h_{\alpha\beta}(z_\beta)$ i.e. of spin-structure $\mathcal{L}$. This Riemann surface with spin-structure, $\Sigma \equiv S^{\text{red}}$, is the *reduced space* of the SRS S . We have an imbedding $\Sigma \rightarrow S$.

The SRS is *split* iff we can choose the coordinates so that $\epsilon_{\alpha\beta} \equiv 0$. In this case S is just a spin-bundle over $\Sigma \equiv S^{\text{red}}$ with odd fibers: we have a projection $S \rightarrow \Sigma$ and the odd coordinate θ is a section of the spin-bundle $\mathcal{L}$. The moduli space $\mathcal{SM}_g^{\text{split}} \subset \mathcal{SM}_g$ of split genus- g SRS is then a 2^{2g} -fold cover of $\mathcal{M}_g$. There is a projection map $\mathcal{SM}_g \rightarrow \mathcal{SM}_g^{\text{split}}$ if and only if the supermanifold $\mathcal{SM}_g$ is *projected*.⁶ This is believed to be false for $g \geq 3$ (and certainly so for $g \geq 5$) [17].

In terms of the fields (g_{ab}, χ_a) , a split SRS is a 2d gravitino χ_a which is globally pure gauge, so can be set to zero. It follows that any SRS S may be seen as its reduced space Σ with a possibly non-trivial gravitino field. Modulo super-Weyl transformations, the gravitino has two components χ and $\tilde{\chi}$. We focus on χ which is

⁶ A supermanifold is *projected* iff we can choose the local coordinates such that local bosonic coordinates m_α^i transform between themselves $m_\alpha^i = f_{\alpha\beta}^i(m_\beta^j)$. Split supermanifolds are in particular projected. See e.g. [17]. For SRS the converse holds: *projected* $\Rightarrow$ *split*.

relevant for the left-moving side of the superstring; the story for $\tilde{\chi}$ is similar. χ is a $(-\frac{1}{2}, 1)$ differential i.e. a smooth section of $\mathcal{L}^{-1} \otimes \bar{K}$. By the usual Hodge argument, the gauge-invariant part of a gravitino field χ is its harmonic projection

$$[\chi] \in H^1(\Sigma, \mathcal{L}^{-1}) \simeq H^0(\Sigma, \mathcal{L}^3)^\vee, \quad \dim_{\mathbb{C}} H^1(\Sigma, \mathcal{L}^{-1}) = 2(g-1). \quad (10.37)$$

Reducing to $\mathcal{M}_g$ Physically it would be preferable to have an expression of the vacuum amplitude as an integral over $\mathcal{SM}_g^{\text{split}}$ (that is, as a sum over spin-structures followed by an integral over $\mathcal{M}_g$) rather than an integral over the supermanifold $\mathcal{SM}_g$ of complex dimension $3(g-1) | 2(g-1)$ even if, geometrically, the representation as an integral over $\mathcal{SM}_g$ is more intrinsic. In other words, one wishes to integrate away the odd moduli to remain with only the even moduli of the underlying Riemann surface Σ . One reason why this is preferable is spacetime SUSY: supersymmetry arises from the GSO projection i.e. from the sum over spin-structures cf. Sect. 3.8. The spin-structure is part of the SRS structure, so spacetime SUSY (if present) is a property of the integral over $\mathcal{SM}_g$ not of the single SRS. If we were able to reduce the amplitude to an integral over $\mathcal{M}_g$, the integrand itself would satisfy the SUSY Ward identities⁷ as we argued in Sect. 3.8 using contour manipulations. The SUSY Ward identities are crucial for most physical considerations, including the perturbative consistency of the theory: we want to keep them as explicit as we can. In order to do that, roughly speaking, we would like to *integrate out* the odd moduli, thus reducing from an integral on $\mathcal{SM}_g$ to an integral over $\mathcal{M}_g$. Unfortunately for $g \geq 3$ there is an obstruction to do this *globally* on $\mathcal{M}_g$ since $\mathcal{SM}_g$ is not projected [17]. One is forced to work locally on $\mathcal{M}_g$, and then patch together the local answers to address the global issues.⁸

Let us be pragmatic. We choose $2(g-1)$ sufficiently generic *commuting* gravitino configurations $\chi_k(z)$; their classes $[\chi_k(z)]$ span $H^1(\Sigma, \mathcal{L}^{-1})$, so that the odd gravitino background $\chi(z, \mu) \equiv \sum_k \mu_k \chi_k(z)$, which depends on $2(g-1)$ *odd* parameters μ_k , gives a complete SUSY gauge-slice i.e. a family of field configurations such that all gravitino field is gauge-equivalent to *precisely one* configuration in the family. This construction is at fixed $\Sigma_{m_0} \equiv S^{\text{red}}$, i.e. for a fixed value $m_0 \in \mathcal{SM}^{\text{split}}$ of the even moduli. The classes $[\chi_k(z)]$ will remain linearly independent in a sufficiently small open neighborhood $U \subset \mathcal{SM}^{\text{split}}$ of m_0 , so we can use $\chi(z, \mu)$ as a gauge-slice in all U . Such a choice of slice is said to be *m-independent* in [2, 6] since $\chi(z, \mu)$ has no explicit dependence on the even moduli m . The use of a *m-independent* slice simplifies the local expression of the integrand in $\mathcal{M}_g$ [2, 6]. However there is a price to pay: the $[\chi_k(z)]$ will cease to be a basis of $H^1(\Sigma_m, \mathcal{L}_m^{-1})$ for some m away from m_0 in $\mathcal{M}_g$. Therefore this choice of gauge-slice, while convenient, applies only locally on $\mathcal{M}_g$, and we have to cover $\mathcal{M}_g$ by sufficiently small open sets, compute the integrands in each set, and compare them on the overlaps.

⁷ The statement is not as innocent as it may sound: there are subtleties with SUSY at higher-genus. We mention them below. See [2, 6].

⁸ On a supermanifold there is a natural volume form only if it is split; otherwise the volume form is ambiguous for a total derivative [4, 6]. This is the main source of technical trouble.

The matter + ghost 2d action in presence of a gravitino background is

$$S = S_{\chi=0} + \int (\chi T_F + \tilde{\chi} \tilde{T}_F) \quad (10.38)$$

where T_F is the total matter + ghost supercurrent

$$T_F = \psi^\mu \partial X_\mu + b\gamma + 3\beta\partial c + (\partial\beta)c. \quad (10.39)$$

The last term in (10.38) is the Noether coupling between the SUSY current T_F and its gauge-field (the gravitino). The left-moving side of the g -loop vacuum amplitude, as a function of the even moduli m^i and odd moduli μ_k , then takes the form

$$\mathcal{Z}(m, \mu) = \int [d(\text{fields})] e^{-S(m, \mu=0)} \prod_i \langle b, h_i \rangle \prod_k (1 + \mu_k \langle \chi_k, T_F \rangle) \delta(\langle \chi_k, \beta \rangle) \quad (10.40)$$

where $\langle \xi, \eta \rangle \equiv \int_\Sigma \xi \eta$ and $\{h_i\}$ is the basis of Beltrami differentials dual to the basis dm^i of $T^*\mathcal{M}_g$. The factor $\prod_i \langle b, h_i \rangle$ is required to absorb the b -ghost zero-modes as in the bosonic string. Likewise the factor $\prod_k \delta(\langle \chi_k, \beta \rangle)$ absorbs the β zero-modes (cf. Sect. 2.5.7). Integrating out the odd moduli μ_k we get

$$\int \prod_k d\mu_k \mathcal{Z}(m, \mu) = \int [d(\text{fields})] e^{-S(m, \mu=0)} \prod_k X[\chi_k] \prod_i \langle b, h_i \rangle \quad (10.41)$$

where

$$X[\chi_k] = \langle \chi_k, T_F \rangle \delta(\langle \chi_k, \beta \rangle) \quad (10.42)$$

are (generalized) “picture changing operators” [6]. The picture changing operators are BRST-invariant but they depend on the largely arbitrary choice of the $2(g-1)$ configurations χ_k . If we choose different gravitino fields $\chi'_k \equiv \chi_k + \varepsilon_k$ the picture changing operators $X[\chi_k]$ change by BRST-exact terms

$$X[\chi_k + \varepsilon_k] - X[\chi_k] = \{Q_{\text{BRST}}, \delta(\langle \chi_k, \beta \rangle) \langle \varepsilon_k, \beta \rangle\}. \quad (10.43)$$

A BRST-exact insertion gives a vanishing amplitude *provided* all other insertions are BRST-invariant. All insertions in (10.41) are BRST-invariant except the b -fields which absorb the ghost zero-modes

$$\{Q_{\text{BRST}}, \langle h_i, b \rangle\} = \langle h_i, T_B \rangle = \frac{\partial}{\partial m^i} S(m, \mu=0), \quad (10.44)$$

so that the variation of the amplitude is a total derivative in $\mathcal{M}_g$ (here it is important that the insertions have no explicit dependence on m). Thus the integrand locally defined in $\mathcal{M}_g$ is ambiguous by a total derivative. To show that this ambiguity is not dangerous is a highly non-trivial issue that we just touch upon.

Comparison with $X(z)$ in Eq.(3.204) It remains to make contact with the local picture-changing operators as defined in Chap.3. The standard choice of the χ_k 's is $\chi_k(z) = \delta(z - z_k)$. For a generic $(2g - 2)$ -tuple of points $\{z_k\} \in \Sigma^{2(g-1)}$ the classes $[\delta(z - z_k)]$ span $H^1(\Sigma, \mathcal{L}^{-1})$. However there is a codimension one space $Y \subset \Sigma^{2(g-1)}$ such that when $\{z_k\} \in Y$ the $[\delta(z - z_k)]$'s fail to span $H^1(\Sigma, \mathcal{L}^{-1})$ [18]. With this standard choice, the picture-changing operators are localized at the points z_k

$$X(z_k) =: \delta(\beta(z_k)) T_F(z_k) =: \{Q_{BRST}, H(\beta(z_k))\}, \quad \{z_k\} \notin Y, \quad (10.45)$$

where $H(x)$ is the Heaviside step function

$$H(x) = \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi i(\omega - i\epsilon)} e^{i\omega x} \equiv \begin{cases} 1 & x > 0 \\ 0 & x \leq 0, \end{cases} \quad \frac{dH(x)}{dx} = \delta(x). \quad (10.46)$$

$X(z)$ has the same form as in Eq.(3.204) for the identifications [2, 6]

$$H(\beta) \leftrightarrow \xi, \quad \delta(\beta) \leftrightarrow e^\phi, \quad \delta(\gamma) \leftrightarrow e^{-\phi}, \quad : \partial \gamma \delta(\gamma) : \leftrightarrow \eta \quad (10.47)$$

with the fields ϕ, ξ, η which bosonize the SUSY ghosts: $\gamma = \eta e^\phi, \beta = e^{-\phi} \partial \xi$.

Exercise 10.3 Show that the identifications (10.47) are consistent with the Virasoro weights, ghost charges, Bose levels, and OPEs.

Recall from Sect. 2.5 that the local operator $e^{\phi(z)}$ increases the Bose-sea level by 1, and hence maps a state in the level- q Bose Fock space $\mathfrak{F}(q) \equiv \mathfrak{A}_{\beta, \gamma}|q\rangle$ into a state in the level $(q + 1)$ Fock space $\mathfrak{F}(q + 1)$, while $\gamma \in \mathfrak{A}_{\beta, \gamma}$ acts inside $\mathfrak{F}(q)$. The OPE

$$\xi(z) \gamma(w) = \xi(z) \eta(w) e^{\phi(w)} \sim \frac{1}{z - w} e^{\phi(w)} \quad (10.48)$$

shows that $\xi(z) \equiv H(\beta(z))$ also maps $\mathfrak{F}(q)$ to $\mathfrak{F}(q + 1)$. The small algebra operator $\rho(z) \equiv \partial \xi(z)$ then increases the sea level by 1. Working in the large algebra we need to insert an extra ‘dumb’ $\xi(z)$ to kill its zero-mode. $\delta(\beta)$ (resp. $\delta(\gamma)$) increases (resp. decreases) the sea level by 1. Hence, to absorb all β zero-modes in the genus- g vacuum amplitude, the numbers of insertions of $\delta(\beta), \delta(\gamma), H(\beta)$ should satisfy [2, 6] (count in the large algebra)

$$2(g - 1) = \text{net variation of Bose-sea level} = \#\delta(\beta) - \#\delta(\gamma) + \#H(\beta) - 1. \quad (10.49)$$

On the other hand, we have the anomalous ghost number conservation (cf. Sect. 2.5)

$$2(g - 1) = \text{net variation of ghost charge} = \#\gamma - \#\beta - \#\delta(\gamma) + \#\delta(\beta) \quad (10.50)$$

Equations (10.49) and (10.50) give the number of insertions for a non-zero g -loop amplitude

$$\#e^\phi - \#e^{-\phi} = 2g - 2, \quad \#\xi - \#\eta = 1 \quad (10.51)$$

The second equation looks puzzling: from the anomaly of the current $\xi\eta(z)$ one would expect the RHS to be $1 - g$ equal to the net number of ξ, η Fermi zero-modes. *What is the origin of this discrepancy?* The point is that the correlation functions of the “distributional” operators $\delta(\beta(z)), \delta(\gamma(z))$ and $H(\beta(z))$ —but not those of the “innocent” local fields $\beta(z), \gamma(z)$ —have, in addition to the physical singularities when z approaches a point $y \in \Sigma$ where a local operator $O(y)$ is inserted, also *g spurious poles* at points where *nothing is inserted*, see **BOX 10.1**. Then, when we perform the standard contour manoeuvres to prove the anomalous Ward identities of the current $\xi\eta(z)$, we pick up—in addition to the usual residues at the physical singularities—*extra* contributions from the residues at the spurious poles which behave as *g “phantom” η insertions*, so that the *apparent* visible variation of the $\xi\eta$ charge is not $1 - g$ but 1 as in (10.51). One checks that the fields $\beta(z), \gamma(z)$ which preserve the Fock spaces $\mathfrak{F}(q)$ have no spurious singularities [2, 6]. This is required by quantum consistency: otherwise we would have spurious contributions to the Ward identities of the BRST current, and this will make the quantization in Chap. 3 meaningless.

BOX 10.1 - Spurious poles in $\delta(\beta), \delta(\gamma), H(\beta)$ correlators

We give a simple explanation of the presence of spurious poles in the correlation of sea-level changing operators $\delta(\beta(z)), \delta(\gamma(z))$. Consider a b, c Fermi system with weight $\lambda > 1/2$ in a $g \geq 2$ surface. The non-zero correlations have the form

$$\left\langle \prod_i^{(2\lambda-1)(g-1)+n} b(z_i) \prod_{a=1}^n c(w_a) \right\rangle \quad n = 0, 1, 2, \dots$$

As a function of z_1 , this correlator has n simple poles at w_a , hence by Riemann–Roch it has $\deg K^\lambda + n \equiv 2\lambda(g-1) + n$ zeros. Fermi statistics predicts zeros at $z_1 = z_i$ for $i \geq 2$: this accounts for $(2\lambda-1)(g-1) + n - 1$ zeros. There remain *g “extra” zeros* at points where *nothing is inserted*. Now replace the Fermi system with the β, γ Bose system with the *same* λ and spin-structure. The path integral is still Gaussian, and all modes (zero and non-zero) are the same ones as for the b, c system—only the statistics gets inverted. Using the rules of Gaussian integration we get the relation (cf. Eq. (10.52) for $\lambda = \frac{3}{2}$)

$$\int [d\beta d\gamma] e^{-S[\beta, \gamma]} \prod_{a=1}^{(2\lambda-1)(g-1)} \delta(\beta(z_a)) = \left[\int [dbdc] e^{-S[b, c]} \prod_{a=1}^{(2\lambda-1)(g-1)} b(z_a) \right]^{-1}$$

The function inside the square bracket has *g “spurious” zeros* in z_1 at points where nothing is inserted. Hence the LHS, as a function of z_1 , has *g spurious poles*. From Eq. (10.49) we see that the result holds also with $\delta(\beta)$ replaced by $H(\beta)$ (after the elimination of the zero-mode of ξ)

Using the Fourier integral representations of the distributions $\delta(x)$ and $H(x)$, we can compute arbitrary correlations of $\xi, \eta, \exp(\pm\phi)$ and ghosts’ spin-fields $\exp[\pm\frac{1}{2}\phi]$, in terms of path integrals over the β, γ system. In turn these are related

to path integrals over the free Fermi system b, c with the same $\lambda = 3/2$ and spin-structure. Examples of relations between the Bose/Fermi path integrals are

$$\int [d\beta d\gamma] e^{-S} \prod_{a=1}^{2g-2} \delta(\beta(z_a)) = \left[\int [dbdc] e^{-S} \prod_{a=1}^{2g-2} b(z_a) \right]^{-1} \quad (10.52)$$

$$\frac{\langle \beta(z) \gamma(z) \prod_{a=1}^{2g-2} \delta(\beta(z_a)) \rangle}{\langle \prod_{a=1}^{2g-2} \delta(\beta(z_a)) \rangle} = - \frac{\langle b(z) c(z) \prod_{a=1}^{2g-2} b(z_a) \rangle}{\langle \prod_{a=1}^{2g-2} b(z_a) \rangle}. \quad (10.53)$$

The explicit expression for the non-zero genus- g amplitude

$$\int [d\xi d\eta d\phi] e^{-S[\xi, \eta, \phi]} \prod_{a=1}^{n+1} \xi(x_a) \prod_{b=1}^n \eta(y_b) \prod_c e^{q_c \phi(z_c)}, \quad (10.54)$$

where the picture-charges q_c are integers or half-integer such that $\sum_c q_c = 2g - 2$, is given in [2, 6], see also [3]. These correlations contain spurious singularities, i.e. poles at points where there is no insertion. The amplitude (10.41)

$$\int [d(\text{fields})] e^{-S(m, \mu=0)} \prod_k X[z_k] \prod_i \langle b, h_i \rangle, \quad (10.55)$$

seen as a function of z_1 , becomes singular at a finite set of points in Σ by effect of the aforementioned spurious singularities.

Note 10.1 The geometric interpretation of the singularities of (10.55) (which appear in codimension 1) is obvious: they correspond to points $z_1 \in \Sigma$ where the gravitino class $[\delta(z - z_1)]$ is not linearly independent of the classes $\{[\delta(z - z_k)]\}_{k=2}^{2g-2}$ —i.e. when $\{z_k\} \in Y \subset \Sigma^{2(g-1)}$ see discussion before Eq. (10.45)—so that the gravitino family $\chi(z, \mu) = \sum_k \mu_k \delta(z - z_k)$ ceases to be a *bona fide* gauge-slice.

In conclusion: working locally in $\mathcal{M}_g$ we can write the amplitude, *after* the integration of odd moduli, as a correlation function of the matter + ghost system on the Riemann surface (with spin-structure) Σ with suitable PCO insertions. The amplitude is ambiguous for total derivatives, and one has to check that these ambiguities are physically harmless: in particular they must not spoil space-time SUSY. Luckily this turns out to be true: see [2, 6] for a proof.⁹ The ambiguity is studied in detail in [4, 19]; in [5] it is shown that it is proportional to the tadpoles of physical massless fields in lower genera. Hence one can fix *unambiguously* the vacuum amplitude recursively in the genus g .

⁹ The statement is referred to superstrings *moving in 10d*. In certain heterotic models with 4 non-compact dimensions with a space-time Abelian gauge symmetry whose anomaly is canceled by a Green–Schwarz counter-term, the loop effects break SUSY spontaneously: for a nice modern treatment, see [10]. This is the only known case of spontaneous breaking of SUSY by loop terms.

Note 10.2 In order to make concrete the PCO approach to higher-genera amplitudes, one needs to construct a convenient open cover of $\mathcal{M}_g$, write the amplitude inside each patch of the cover in terms of pertinent PCO insertions, and then find the appropriate gluing of the local expressions in the overlaps between patches. An explicit such procedure is described in detail in [18].

Adding Punctures In a SRS we have two kinds of punctures: NS and R. Besides the punctures may carry different pictures. We focus on the standard pictures (-1) for NS and $(-\frac{1}{2})$ for R. The gravitino has a pole at an ordinary NS puncture, while ordinary R punctures are in even number, and they create a square-root cut in the gravitino field χ . Two R punctures reckon as a NS puncture for dimension counting: the gravitino field χ is a smooth section $(\mathcal{L}^{-1} \otimes L) \otimes \overline{K}$ where L is a holomorphic line bundle of degree $-n_{\text{NS}} - n_{\text{R}}/2$ where n_{NS} (n_{R}) is the number of NS (resp. R) punctures (all in standard picture). Then the number of odd moduli is (n_{R} is even)

$$\dim H^1(\Sigma, \mathcal{L}^{-1} \otimes L) = 2g - 2 + n_{\text{NS}} + \frac{n_{\text{R}}}{2} \quad \text{for } g \geq 2, \quad (10.56)$$

If we use only standard picture vertices, we have to insert $n_X \equiv \dim H^1(\Sigma, \mathcal{L}^{-1} \otimes L)$ PCOs. We can use some of these to convert the NS vertices to 0 picture and *half* of the R vertices to $+\frac{1}{2}$ picture, remaining with $\dim H^1(\Sigma, \mathcal{L}^{-1})$ explicit PCO insertions as for the vacuum amplitude.

Note 10.3 There are several other important issues in higher-loop amplitudes which we do not address, in particular mass-renormalization and related phenomena. The interested reader may have a look to [20–27].

10.3 One-Loop Amplitudes

The subtleties mentioned in Sect. 10.2 play no role at $g = 1$ where a simply-minded PCO approach is fully justified. In this section we give a couple of important examples of one-loop amplitudes and discuss their physical implications.

On the torus in an even spin-structure ($\mathcal{L} \not\simeq \mathcal{O}$) the ghosts β, γ have no zero-modes, and for each $m \in \mathcal{M}_1$ the *only* SRS with spin-structure $\mathcal{L}$ is the split one. In the odd spin-structure (P, P) ($\mathcal{L} \simeq \mathcal{O}$) both ghosts β, γ have one zero-mode which is constant since the torus is an Abelian group, hence homogeneous. The gravitino class $[\delta(z - z_0)]$ is independent of z_0 , hence there are no spurious poles (cf. Note 10.1), and the PCO formalism yields a *global* representation of the amplitude as an integral over $\mathcal{M}_1$. This is consistent since the supermoduli $\mathcal{SM}_1$ is split.

Heterotic String: Four Gauge Bosons, One 2-Form

The Green–Schwarz mechanism (cf. Sect. 9.2) requires a one-loop Chern–Simons term

$$\propto \int B^{(2)} \text{tr } F^4. \quad (10.57)$$

We wish to confirm the presence of this coupling by an explicit string computation. It arises from the (P, P) sector of the torus path integral since the coupling (10.57) is *odd* under space-time parity: written in components it involves a 10d ϵ -tensor. The heterotic string action and its constraints are invariant under space-time parity, and the parity asymmetry of the theory arises from the GSO projection in the right-moving R sector where we take $(-1)^{\tilde{F}}$ to be $+1$ instead of -1 so that the massless fermions are $\mathbf{16}_s$ spinors instead of $\mathbf{16}_c$ ones. Then (P, P) is the only spin-structure whose path integral is odd under space-time parity: it is the trace in the R sector with $(-1)^{\tilde{F}}$ inserted, and so it changes sign under $(-1)^{\tilde{F}} \leftrightarrow -(-1)^{\tilde{F}}$.

The b, c ghosts work as in the bosonic case, and we have a factor $1/2$ from the GSO projection. In the (P, P) spin-structure we need one $\delta(\tilde{\gamma})$ to absorb the $\tilde{\gamma}$ zero-mode, and one PCO insertion

$$\tilde{X}(\bar{z}) = \tilde{T}_F(\bar{z}) \delta(\tilde{\beta}(\bar{z})) - \partial \tilde{b}(\bar{z}) \delta'(\tilde{\beta}(\bar{z})) \quad (10.58)$$

from the $\tilde{\beta}$ zero-mode. The other 4 PCO predicted by (10.56) are used to convert 4 vertices to 0 picture, leaving the last vertex in picture -1 . The amplitude is then¹⁰

$$\begin{aligned} & \left(\frac{2}{\alpha'} \right)^{5/2} g_c^5 \int_{F_0} \frac{d\tau d\bar{\tau}}{8\tau_2} \left[\prod_{i=1}^5 \int d^2 w_i \right] \left\langle b(0) \tilde{b}(0) c(0) \tilde{c}(0) \tilde{X}(0) \times \right. \\ & \times \left[\prod_{i=1}^4 \hat{k}^{-1/2} j^{a_i} (i e_i \cdot \bar{\partial} X + \frac{1}{2} \alpha' k_i \cdot \tilde{\psi} e_i \cdot \tilde{\psi}) e^{i l_i \cdot X}(w_i, \bar{w}_i) \right] \times \\ & \times \left. i e_{5\mu\nu} \partial X^\mu \delta(\tilde{\gamma}) \tilde{\psi}^\nu e^{i k_5 \cdot X}(w_5, \bar{w}_5) \right\rangle_{(P,P)} \end{aligned} \quad (10.59)$$

We consider separately the path integrals for the $\tilde{\psi}^\mu$, X^μ , b, c , $\tilde{\beta}$, $\tilde{\gamma}$ systems and for the j^a gauge current algebra. In the (P, P) b.c. the integral over $\tilde{\psi}^\mu$ vanishes by zero-mode counting unless there are at least ten factors of $\tilde{\psi}^\mu$ inserted. In the amplitude (10.59) we have a maximum of ten $\tilde{\psi}^\mu$ including one from the term

$$\delta(\tilde{\beta}) (2/\alpha')^{1/2} \tilde{\psi}^\rho \bar{\partial} X_\rho \quad (10.60)$$

in the PCO operator $\tilde{X}$. The $\tilde{\psi}^\mu$ path integral is easily computed as a Hilbert trace

$$\begin{aligned} \left\langle \prod_{i=1}^{10} \tilde{\psi}^{\mu_i} \right\rangle_{\tilde{\psi} (P,P)} &= \epsilon^{\mu_1 \dots \mu_{10}} \bar{q}^{h-c/24} \prod_{n=1}^{\infty} (1 - \bar{q}^n)^{10} = \\ &= \epsilon^{\mu_1 \dots \mu_{10}} \bar{q}^{10/24} \prod_{n=1}^{\infty} (1 - \bar{q}^n)^{10} = \epsilon^{\mu_1 \dots \mu_{10}} [\eta(\tau)^{10}]^*, \end{aligned} \quad (10.61)$$

¹⁰ As always, F_0 is the fundamental domain of the modular group, cf. Fig. 4.3.

where $h = \frac{5}{8}$ is the weight of a $SO(10)$ spin field. The X^μ path integral is reduced to

$$\left\langle \partial X^\mu(w_5) \bar{\partial} X^\rho(0) \prod_{i=1}^5 e^{ik_i \cdot X(w_i, \bar{w}_i)} \right\rangle_X, \quad (10.62)$$

where the derivatives come from the 2-form vertex and the PCO. To check the presence of the counter-term (10.57) it suffices to consider the limit $k_i \rightarrow 0$. In this limit the contractions between gradients and exponentials and between exponentials are suppressed. Only the contraction between the gradients survives, giving

$$-\frac{\alpha'}{8\pi\tau_2} \quad (10.63)$$

from the term $-\alpha'(\text{Im } w_{ij})^2/4\pi\tau_2$ in the torus Green function, see Eq. (4.39). The low-momentum limit of the amplitude (10.62) is then

$$-i(2\pi)^{10} \delta\left(\sum_i k_i\right) \frac{\eta^{\mu\nu} \alpha'}{8\pi\tau_2 (4\pi^2 \alpha' \tau_2)^5 |\eta(\tau)|^{20}} \quad (10.64)$$

The b, c path integral yields

$$\langle b(0) \tilde{b}(0) \tilde{c}(0) c(0) \rangle_{b,c} = |\eta(\tau)|^4, \quad (10.65)$$

as in the bosonic string. Finally we need the 4-current function

$$\hat{k}^{-2} \langle j^{a_1}(w_1) j^{a_2}(w_2) j^{a_3}(w_3) j^{a_4}(w_4) \rangle_{\text{current algebra}}. \quad (10.66)$$

In our conventions $\hat{k} = 1/2$. In the limit $k_i \rightarrow 0$ Eq. (10.66) is the only factor in the integrand of (10.59) which depends on the w_i 's. The integration over the position has the effect of averaging over $\text{Re } w_i$ and so *naively* we may replace each current $j^{a_i}(w_i)$ with the corresponding charge Q^{a_i} , and then evaluate the path integral as a trace over the Hilbert space. This is not quite correct because of a contact-term when two currents come together

$$j^a(w) j^b(0) \Big|_{\text{contact term}} = -\pi \delta(w, \bar{w}) \delta^{ab} \equiv -\frac{1}{8\pi\tau_2} + \text{total derivative}. \quad (10.67)$$

The simplest way to get this subtle term is to compare the generating function for the currents j^a , integrated along the A -cycle, with the Hilbert space trace with the insertion of fugacities y^a for the charges $Q^a = \oint_A j^a$. We claim that

$$\left\langle \exp\left(y^a \oint_A j^a\right) \right\rangle_{\text{torus } \tau} = \exp\left[-\frac{y \cdot y}{16\pi\tau_2}\right] \text{Tr}\left[e^{2\pi i \tau (L_0 - c/24)} e^{y \cdot Q}\right], \quad (10.68)$$

see **BOX 10.2**. The Gaussian prefactor arises from the contact term (10.67).

The trace in (10.68) is easily performed in the bosonic formulation of the heterotic string, where it reduces to a sum over the $\text{Spin}(32)^+$ or $E_8 \times E_8$ self-dual lattice Γ

$$f(y, q) \stackrel{\text{def}}{=} \left\langle \exp \left(y \cdot \oint j \right) \right\rangle = \frac{e^{-y \cdot y / (16\pi\tau_2)}}{\eta(q)^{16}} \sum_{\ell \in \Gamma} q^{\ell^2/2} \exp(y \cdot \ell / \sqrt{2}). \quad (10.69)$$

Gathering all factors, including $(8\pi\tau_2)^5$ from the integration¹¹ over the d^2w_i , in the limit $k_i \rightarrow 0$ the amplitude reduces to

BOX 10.2 - The generating function of integrated current correlators

We show Eq. (10.68). For simplicity we take $G = SU(2)$ at level 1 and bosonize it in terms of a periodic scalar ϕ at self-dual radius $R^2 = \alpha' = 2$. Then

$$\left\langle \exp \left(\frac{y}{2\pi} \oint_A j + \frac{\tilde{y}}{2\pi} \oint_A \tilde{j} \right) \right\rangle = \int [d\phi] \exp \left[-\frac{1}{4\pi} \int \partial\phi \bar{\partial}\phi + \oint_A \frac{y \partial\phi + \tilde{y} \bar{\partial}\phi}{2\sqrt{2}\pi} \right] \spadesuit$$

As in Sect. 6.2 (cf. Eqs. (6.52)–(6.57)) the path integral over the periodic scalar, $\phi \sim \phi + 2\sqrt{2}\pi$, decomposes into several topological sectors labeled by two integers m, w . In the sector m, w we write (cf. (6.56))

$$\partial\phi(z) = i \frac{w\bar{\tau} - m}{\sqrt{2}\tau_2} dz + \partial\varphi(z), \quad \bar{\partial}\phi(\bar{z}) = -i \frac{w\tau - m}{\sqrt{2}\tau_2} d\bar{z} + \bar{\partial}\varphi(\bar{z})$$

where φ is a scalar field which is univalued on the torus. The path integral ($\spadesuit$) factorizes into a ‘classical’ sum over the topological sectors times a path integral for the non-compact scalar φ . The second factor was computed in Eq. (4.9)

$$\mathbf{det}'(-\bar{\partial}\partial)^{-1/2} \exp \left[-\frac{1}{16\pi^2} \oint_A dz \oint_A dw (y\partial_z + \tilde{y}\partial_{\bar{z}})(y\partial_w + \tilde{y}\partial_{\bar{w}}) G(z, w) \right]$$

where $G(z, w)$ is the scalar Green’s function on the torus given by Eq. (4.39) with $\alpha' = 2$. $\mathbf{det}'(-\bar{\partial}\partial)^{-1/2} = |\eta(\tau)|^{-2}$ while the exponential factor is

$$\exp \left[-\frac{(y - \tilde{y})^2}{16\pi\tau_2} \right]$$

Setting $\tilde{y} = 0$ and taking only the left-moving (holomorphic) factor, we get

$$\left\langle \exp \left(\frac{y}{2\pi} \oint_A j \right) \right\rangle = \left(\frac{1}{\eta(q)} \sum_{P_L} q^{P_L^2/2} e^{y P_L / \sqrt{2}} \right) \exp \left[-\frac{y^2}{16\pi\tau_2} \right] \equiv \text{Tr} [q^{L_0 - c/24} e^{zQ}] \cdot \exp \left[-\frac{y^2}{16\pi\tau_2} \right]$$

¹¹ This is just $(\int_{\text{torus}} d^2w)^5$ where the torus has periods $(2\pi, 2\pi\tau)$ and the measure $d^2w \equiv 2 dx dy$.

$$\begin{aligned}
& -\frac{i g_c}{\pi(\alpha')^3} (2\pi)^{10} \delta\left(\sum_i k_i\right) \times \\
& \times \epsilon^{\mu_1 \mu_2 \mu_3 \mu_4 \mu_5 \mu_6 \mu_7 \mu_8 \mu_9 \mu_{10}} k_{1\mu_1} e_{1\mu_2} k_{2\mu_3} e_{2\mu_4} k_{3\mu_5} e_{3\mu_6} k_{4\mu_7} e_{4\mu_8} e_{5\mu_9 \mu_{10}} \times \\
& \times \int_{F_0} \frac{d^2 \tau}{\tau_2^2} \frac{\partial^4 U(y, q)}{\partial y^{a_1} \partial y^{a_2} \partial y^{a_3} \partial y^{a_4}} \Big|_{z=0},
\end{aligned} \tag{10.70}$$

where

$$U(y, \tau) \stackrel{\text{def}}{=} \eta(q)^{-8} f(y, e^{2\pi i \tau}). \tag{10.71}$$

Making the replacements

$$e_{\mu\nu} \rightarrow B_{\mu\nu}/2\kappa, \quad k_{[\mu} e_{\nu]} = -i F_{\mu\nu}/2g_{\text{YM}}, \tag{10.72}$$

converting from the tensor to the form notation, and using the relations between the couplings, the amplitude may be written as the effective action interaction

$$-\frac{1}{2^9 \pi^6 \alpha'} \int_M B^{(2)} \int_{F_0} \frac{d^2 \tau}{\tau_2^2} U(F^{(2)}, \tau). \tag{10.73}$$

The integration over the space-time M picks up the degree-8 part $U(F^{(2)}, \tau)^{(8)}$ of the even form $U(F^{(2)}, \tau)$. To understand the modular properties of the integrand in (10.73), we compare it with the analysis of anomaly cancelation in Sect. 9.3; as explained there, the heterotic string propagates consistently in a space-time with $R^{(2)} = 0$ only if $[\text{tr}(F^{(2)})^2] = 0$. In such a background

$$[U(F^{(2)}, \tau)] = [T(\tau, F^{(2)})] = [A(\tau, R = 0, F^{(2)})] \quad \text{iff } [\text{tr}(F^{(2)})^2] = 0. \tag{10.74}$$

Hence from Eq. (9.86) with $n \equiv (d-2)/2 = 4$

$$U\left(\frac{F^{(2)}}{c\tau + d}, \frac{a\tau + b}{c\tau + d}\right) = (c\tau + d)^{-4} U(F^{(2)}, \tau), \tag{10.75}$$

$$U\left(F^{(2)}, \frac{a\tau + b}{c\tau + d}\right)^{(8+2k)} = (c\tau + d)^k U(F^{(2)}, \tau)^{(8+2k)} \tag{10.76}$$

in a consistent background. Thus (10.73) is modular invariant provided we restrict the gauge fields $F^{(2)}$ to the physically allowed ones.

To simplify the integral over the fundamental domain F_0 we note that

$$\partial_{\bar{\tau}} U(y, \tau) \equiv U(y, \tau) \partial_{\bar{\tau}} \left(\frac{-i y \cdot y}{8\pi(\tau - \bar{\tau})} \right) = -\frac{y \cdot y}{32\pi i} \frac{U(y, \tau)}{\tau_2^2}. \tag{10.77}$$

Consider the expression

$$i \frac{U(y, \tau)^{(8)}}{\tau_2^2} d\tau \wedge d\bar{\tau} = -\frac{32\pi}{y \cdot y} d(U(y, q)^{(12)} d\tau), \quad (10.78)$$

from (10.76) we see that the LHS (hence the RHS), seen as a 2-form in the upper half-plane $\mathcal{H}$, becomes modular invariant in the limit $y \cdot y \rightarrow 0$. Then, modulo terms which vanish as $y \cdot y \rightarrow 0$, we have

$$i \int_{F_0} d\tau \wedge d\bar{\tau} \frac{U^{(4)}}{\tau_2^2} = -\frac{32\pi}{z \cdot z} \int_{\partial F_0} U(y, \tau)^{(12)} d\tau = -\frac{32\pi}{y \cdot y} U(y, i\infty)^{(12)}, \quad (10.79)$$

since, by modular invariance, only the boundary at $\tau = i\infty$ gives a non-zero contribution. From Eqs. (10.69) and (10.71), and the definition of the η -function $\eta(q) = q^{1/24}(1 + O(q))$, we see that

$$U(y, \tau)^{(12)} = q^{-24/24}(1 + O(q)) \sum_{\ell \in \Gamma} q^{\ell^2/2} (e^{y \cdot \ell / \sqrt{2}})^{(12)}, \quad (10.80)$$

so that only lattice momenta with $\ell^2 = 2$ contribute to the $q \rightarrow 0$ limit: these are the roots of the gauge group G , and the lattice sum reduces to a trace in the adjoint representation.¹² For $G = SO(32)$ or $E_8 \times E_8$, the trace identities of Sect. 9.2 yield

$$U(y, i\infty)^{(12)} \equiv \text{tr}_{\text{ad}} y^6 \propto \text{tr}(y^2) X(y)^{(8)} \quad (10.81)$$

with $X(y)^{(8)}$ the polynomial (cf. Eq. (9.44))

$$X(y)^{(8)} = \frac{1}{3} \text{tr}_{\text{ad}} y^4 - \frac{1}{900} (\text{tr}_{\text{ad}} y^2)^2, \quad (10.82)$$

so that

$$\frac{U(y, i\infty)^{(12)}}{y \cdot y} \propto X(y)^{(8)}, \quad (10.83)$$

and the effective coupling (10.73) is proportional to

$$\int B^{(2)} \wedge X(F)^{(8)}, \quad (10.84)$$

as required by the anomaly cancelation.

¹² The fugacities y^a belong to the Cartan subalgebra of G , so the traces over the adjoint is a sum over the roots.

Exercise 10.4 Show: overall coefficient agrees with the Green-Schwarz mechanism.

Similarly one sees that the counter-terms required to cancel the gravitational and mixed anomalies are present with the right coefficients. This is hardly a surprise: as the reader doubtlessly has already noticed, the computation of the leading low-momenta terms in the relevant one-loop amplitudes is a rephrasing of the conceptual arguments in Sect. 9.3 and the related theory of the elliptic genus. Thus those arguments, not only guarantee that the anomaly polynomial factorizes, but also that the expected counter-terms are present. For more details, see [28].

Heterotic String: Four Massless Bosons at One-Loop

Next we consider the one-loop four massless bosons amplitudes. Our goal is to describe the structure of the computation and to provide a direct proof of the (general) non-renormalization theorems inferred on physical grounds in Chap. 8. This amplitude gets contributions only from the three even spin-structures; indeed, we have one less vertex than in (10.59) and there are not enough $\tilde{\psi}^\mu$ insertions to soak up the Fermi zero-modes of the (P, P) integral. The amplitude is then

$$\frac{4g_c^4}{(\alpha')^2} \int_F \frac{d\tau d\bar{\tau}}{8\tau_2} \left[\prod_{i=1}^4 d^2 w_i \right] \sum_{\gamma \neq (P, P)} \left\langle b(0) \tilde{b}(0) \tilde{c}(0) c(0) \times \right. \\ \left. \times \left[\prod_{i=1}^4 \hat{k}^{-1/2} j^{a_i} (i e_i \cdot \bar{\partial} X + \frac{1}{2} \alpha' k_i \cdot \tilde{\psi} e_i \cdot \tilde{\psi}) e^{i k_i \cdot X}(w_i, \bar{w}_i) \right] \right\rangle_\gamma. \quad (10.85)$$

For the even spin-structures all vertices are in the 0 picture and there is no PCO, see **BOX 10.3**. This kind of computation simplifies in the light-cone superstring formalism and we shall effectively convert the calculation to that set-up. We first analytically continue the *on-shell* momenta to complex values such that $k^0 = k^1 = 0$. Also we take the polarizations to vanish in the longitudinal directions. The longitudinal degrees of freedom then do not appear in the vertex operators, and the path integrals over longitudinal d.o.f. in the (P, A) , (A, P) and (A, A) sectors just give determinants which cancel against the ghost path integrals. In particular, the combined longitudinal and ghost path integrals in these sectors are independent of the spin-structure, so the net spin-structure dependence comes only from the eight transverse $\tilde{\psi}^i$. We remain with a torus path integral over the transverse d.o.f. summed over the three even spin-structures $(\alpha, \beta) \neq (1, 1)$. From Eq. (5.101) we have $(-1)^{F_{\text{GSO}}} = -(-1)^{F_{\text{trans}}}$ and the ghosts leave beyond the usual signs $(-1)^{\alpha+\beta}$ in the sum over the transverse spin-structures just as for the vacuum amplitude (5.102).

BOX 10.3 - Spin-structures and pictures in Eq. (10.85)

We explain the spin-structure/picture charge assignments in the amplitude (10.85). To soak up the zero-modes, we have to insert one $\delta(\tilde{\gamma})$ per $\tilde{\gamma}$ zero-mode and one $\delta(\tilde{\beta})$ per $\tilde{\beta}$ zero-mode. A picture -1 vertex introduces a $\delta(\tilde{\gamma})$ and a PCO a $\delta(\tilde{\beta})$. In an amplitude at given genus and spin-structure, with only NS vertex operators, the most natural way to write the amplitude is with as many vertices in the -1 picture as $\tilde{\gamma}$ zero-modes all the other vertices in the 0 picture, and the insertion of as many PCOs as zero-modes of $\tilde{\beta}$. The Riemann–Roch theorem says

$$\#(\tilde{\beta} \text{ zero modes}) - \#(\tilde{\gamma} \text{ zero modes}) = 2g - 2,$$

independently of the spin-structure but the actual number of $\tilde{\beta}$ and $\tilde{\gamma}$ zero-modes depends on the particular spin-structure. Recall that a spin-structure is a holomorphic line bundle $\mathcal{L}$ such that $\mathcal{L}^2 = K$ (the canonical bundle) and

$$\#(\tilde{\gamma} \text{ zero modes}) = \dim H^0(\Sigma, \mathcal{L}^{-1}),$$

$$\#(\tilde{\beta} \text{ zero modes}) = \dim H^0(\Sigma, K \otimes \mathcal{L}) \equiv \dim H^1(\Sigma, \mathcal{L}^{-1}) \equiv \dim H^0(\Sigma, \mathcal{L}^{-1}) + 2 - 2g$$

where we used Serre duality and the Riemann–Roch theorem. If $g > 1$, $\deg \mathcal{L}^{-1} = 1 - g < 0$ and $\mathcal{L}^{-1}$ cannot correspond to an effective divisor, so $\dim H^0(\Sigma, \mathcal{L}^{-1}) = 0$, hence that there are no $\tilde{\gamma}$ zero modes and $(2g - 2) \tilde{\beta}$ zero-modes independently of the spin-structure $\mathcal{L}$. However $g = 1$ is special. In this case $K \simeq \mathcal{O}$ is the trivial line bundle, and $\mathcal{L}$ is then a 2-torsion line bundle corresponding to the 2-torsion points in the Jacobian, which is the torus itself. There two kinds of 2-torsion points: the origin corresponding to the trivial line bundle (i.e. the (P, P) spin-structure) and the three half-period points $1/2, \tau/2$ and $(1 + \tau)/2$ which correspond to non-trivial line bundles of degree 0. For a degree zero line bundle $\mathcal{L}$ we have $\dim H^0(\Sigma, \mathcal{L}^{-1}) = 1$ if $\mathcal{L}$ is trivial and zero otherwise. Thus for $g = 1$ in the (P, P) spin-structure we have one $\tilde{\gamma}$ and one $\tilde{\beta}$ zero-modes, while in the other spin-structures we have no $\tilde{\beta}, \tilde{\gamma}$ zero-modes.

The Re-Fermionization Trick To compute the amplitude we use a handy trick based on the re-fermionization of the $Spin(8)$ MW fermions $\tilde{\psi}^i$, see Sect. 2.9.3, and the related Riemann identities in Sect. 6.3. First we add to the transverse d.o.f. path integral the (P, P) sector, even if in the actual amplitude this term is multiplied by zero because of the zero-modes of $\tilde{\psi}^{0,1}$. The sum $\frac{1}{2} \sum_{(\alpha, \beta)} (-1)^{\alpha+\beta} \langle \dots \rangle_{(\alpha, \beta)}$ over the 4 transverse spin-structures $(\alpha, \beta) \in \mathbb{Z}_2^2$ produces a GSO projection of the transverse $Spin(8)^+$ current algebra generated by $\tilde{\psi}^i$. We know from Sect. 2.9.3 that the corresponding $Spin(8)^+$ spin-fields, $\tilde{S}_\alpha$ ($\alpha = 1, \dots, 8$) are eight free MW fermions (the re-fermionization of the bosons which bosonize the $\tilde{\psi}^i$). The spin-structures for the two sets of free MW fermions, the $\tilde{\psi}^i$'s and the $\tilde{\theta}_\alpha$'s, are related as follows

$$\frac{1}{2} \sum_{(\alpha, \beta) \in \mathbb{Z}_2^2} (-1)^{\alpha+\beta} \langle \dots \rangle_{\tilde{\psi}, (\alpha, \beta)} = \langle \dots \rangle_{\tilde{S}_\alpha, (P, P)} \quad (10.86)$$

which is Riemann's identity (6.93) for $\alpha = \beta = 1$. Let us give a more physical argument for the amazing identity (10.86): the spin-fields $\tilde{S}_\alpha$ survive the GSO projection,

which means that after the sum over the spin-structures they become single-valued on the torus, hence periodic along all cycles.

We can compute the amplitude as a (P, P) path integral over the free fermions $\tilde{S}_\alpha$ instead of a sum of path integrals over the original fields $\tilde{\psi}^\mu$. To do this we must rewrite the inserted vertices in the refermionized form. This is easy: equating the $\text{spin}(8)$ currents written in terms of the two sets of Fermi d.o.f. we have

$$\tilde{\psi}^{[i} \tilde{\psi}^{j]} = \frac{1}{4} \tilde{S}^t \Gamma^{ij} \tilde{S}, \quad (10.87)$$

where Γ^{ij} are $\text{Spin}(8)$ gamma-matrices.

The free fermions $\tilde{S}_\alpha$ in the (P, P) structure have 8 zero-modes, and so there must be 8 insertions of $\tilde{S}_\alpha$ to get a non-zero result. The fermions in the vertex operators provide precisely the eight $\tilde{S}_\alpha$ needed to soak up the zero-modes, and we get

$$\begin{aligned} \left\langle \prod_{a=1}^4 \frac{1}{4} \tilde{S}^t \Gamma^{i_a j_a} \tilde{S} \right\rangle_{\tilde{\theta}, (P, P)} &= \frac{1}{2^8} \epsilon^{\alpha_1 \alpha_2 \alpha_3 \alpha_4 \alpha_5 \alpha_6 \alpha_7 \alpha_8} \Gamma_{\alpha_1 \alpha_2}^{i_1 j_1} \dots \Gamma_{\alpha_7 \alpha_8}^{i_4 j_4} = \\ &= t^{i_1 j_1 \dots i_4 j_4} + \epsilon^{i_1 j_1 \dots i_4 j_4}, \end{aligned} \quad (10.88)$$

where $t^{i_1 j_1 \dots i_4 j_4}$ is the tensor which appears in the tree-level four-boson amplitude.

It remains to separate out the spurious contribution of the (P, P) path integral over $\tilde{\psi}^i$: this is the only term which is odd under space-time parity, so it is responsible for the ϵ -term in the RHS of (10.88). Thus we omit this term which does not contribute anyhow, since it is contracted with momenta which are not linearly independent. Further the $SO(8)$ tensor t has a unique $SO(9, 1)$ covariant extension $t^{\mu\nu\sigma\rho\alpha\beta\gamma\delta}$ defined by following identity valid for all antisymmetric tensors $T_{\mu\nu}$

$$t^{\mu\nu\sigma\rho\alpha\beta\gamma\delta} T_{\mu\nu} T_{\sigma\rho} T_{\alpha\beta} T_{\gamma\delta} = \frac{3}{8} \left(4 T_{\mu\nu} T^{\nu\sigma} T_{\sigma\rho} T^{\rho\mu} - (T_{\mu\nu} T^{\nu\mu})^2 \right). \quad (10.89)$$

For $G = SO(32)$ the one-loop correction to the 4-vector effective coupling is

$$\frac{1}{2^8 \pi^5 4! \alpha'} t^{\mu\nu\sigma\rho\alpha\beta\gamma\delta} \text{tr}(F_{\mu\nu} F_{\sigma\rho} F_{\alpha\beta} F_{\gamma\delta}). \quad (10.90)$$

Exercise 10.5 Write the corresponding expression for $G = E_8 \times E_8$.

Note 10.4 The 2d fields $\tilde{S}_\alpha$ carrying a space-time spinor index are the fundamental 2d d.o.f. in the Green–Schwarz formulation [29]. They reduce to free fermions in the light-cone gauge, but are interacting field in a covariant gauge.

10.4 Non-Renormalization Theorems Again

The computation in the previous section implies *inter alia* that any one-loop amplitude with three or fewer massless particles vanishes because there are not enough insertions of $\tilde{S}_\alpha$ to soak up all zero-modes. Since the Newton constant κ^2 can be measured in the three-graviton process, we conclude that it is not renormalized. Moreover all amplitudes vanish at least as k^4 when $k \rightarrow 0$ from the explicit momentum factors in the vertices. This has the important consequence that a constant background

$$G_{\mu\nu}(x) = \eta_{\mu\nu}, \quad \Phi(x) = \Phi_0, \quad (10.91)$$

remains a solution to the field equations to one-loop order. No scalar potential

$$\int d^{10}x \sqrt{-G} V(\Phi), \quad (10.92)$$

is generated. Actually we already know this from the computation of the one-loop vacuum energy in Chap. 5. The absence of correction of the form (10.92) may also be understood in terms of spacetime supersymmetry, see Chap. 8. Using this viewpoint, we conclude that the result is exact, even at the non perturbative level *provided* SUSY is not anomalous, see [30] for more details.

Non-Renormalization Theorems to All Loop-Orders

The one-loop non-renormalization theorem for 0, 1, 2, and 3-point massless amplitudes—which, in particular, implies absence of massless tadpoles—may be extended to all loop-orders in superstring perturbation theory. Formally the proof is a straightforward generalization of the one-loop argument. However, technically we must be sure that the SUSY Ward identities hold at the higher-loop level. This is not obvious due to the ambiguities of the induced measure on $\mathcal{M}_g$, cf. Sect. 10.2. Luckily, the validity of the SUSY Ward identities is guaranteed by the careful analysis of [2, 6] and [5]; then we can safely use the formal argument [31] that we now outline.

As in the previous section we focus on the heterotic string case. We consider the path integral with $n \leq 3$ massless insertions on a given genus $g \geq 1$ surface Σ . We select a non-separating, homotopically non-trivial, simple, closed curve $L \subset \Sigma$, and cut the surface along L to get an open surface Σ' with two boundary components L' and L'' of opposite orientation, see Fig. 10.1. As in the discussion of the state/operator



Fig. 10.1 A genus $g = 3$ surface Σ cut along a non-separating curve $L \subset \Sigma$ leaving a $g = 2$ open surface with boundaries L', L''

correspondence in Sect. 2.2.2, the path integral in this open surface, with Dirichlet boundary conditions, produces a matrix element of the general form

$$\langle \Psi' | \mathcal{U} | \Psi'' \rangle \quad (10.93)$$

where the states $|\Psi'\rangle, |\Psi''\rangle$ are NS/R string states defined on the boundary curves $L' \simeq S^1$ and $L'' \simeq S^1$, respectively. The operator $\mathcal{U}$ depends on the geometry of the open surface Σ' and the vertices inserted in it. The g -loop amplitude is then the *supertrace*

$$\sum_{\Psi \in \text{NS}} \langle \Psi | \mathcal{U} | \Psi \rangle - \sum_{\Psi \in \text{R}} \langle \Psi | \mathcal{U} | \Psi \rangle \equiv \text{Str } \mathcal{U}, \quad (10.94)$$

where the minus sign in the second term arises from the fact that the Ramond states are spacetime fermions. Now suppose that $\mathcal{U}$ may be decomposed into a sum $\sum_a \mathcal{U}_a$ such that for each operator $\mathcal{U}_a$ there is a Hermitian spacetime supercharge $\tilde{Q}_{(a)}$ commuting with it, $[\tilde{Q}_{(a)}, \mathcal{U}_a] = 0$. Since $|\Psi\rangle_{\text{NS}} \rightarrow (\tilde{Q}_{(a)}|\Psi\rangle)_{\text{R}}$ is an isomorphism between the NS and R sectors, we get $\text{Str } \mathcal{U}_a = 0$ and hence $\text{Str } \mathcal{U} = 0$. $\tilde{Q}_{(a)}$ may be written as a contour integral along ℓ' of an anti-holomorphic one-form

$$\tilde{Q}_{(a)}(\bar{z}) = u_{(a)}^\alpha \tilde{S}(\bar{z})_\alpha e^{-\tilde{\phi}(\bar{z})}, \quad \tilde{S}(\bar{z})_\alpha \text{ a Spin}(9, 1) \text{ spin-field} \quad (10.95)$$

which is globally defined (i.e. univalued) in Σ' as a consequence of the GSO projection, cf. Sects. 3.8 and 7.2. This is the point where we use the analysis of [2, 6] to justify our formal manipulations. The $\text{Spin}(9, 1)$ spin-field $\tilde{S}_\alpha(\bar{z})$ can be written as $\tilde{S}_a^{\text{tr}} e^{\pm \frac{i}{2} \tilde{\varphi}(\bar{z})}$ where $\tilde{S}_a^{\text{tr}}(\bar{z})$ is a transverse $SO(8)$ spin-field ($a = 1, \dots, 8$) and $\tilde{\varphi}$ is the scalar which bosonizes the longitudinal current $\tilde{\psi}^0 \tilde{\psi}^1(\bar{z})$.

In absence of vertex insertions $\mathcal{U}$ commutes with all supercharges just because we may deform the contour from ℓ' to the homologous contour ℓ'' and then $\text{Str } \mathcal{U} = 0$. The conclusion remains valid in presence of insertions provided there is *at least one* supercharge commuting with all insertions. The $\tilde{S}_a^{\text{tr}}$'s transform in the vector representation of a triality-rotated $SO(8)$ which acts on a $\mathbb{R}^8$ with orthonormal basis v_a . We have at most $2n$ insertions of $\tilde{S}_a^{\text{tr}}$. The operator $\mathcal{U}$ is a sum of terms

$$\mathcal{U} = \sum_{s=0}^{2n} \sum_{a_i=1}^8 \mathcal{U}_{(a_1, \dots, a_s)} \quad (10.96)$$

where $\mathcal{U}_{(a_1, \dots, a_s)}$ is defined by the path integral over Σ' with $s \leq 2n$ transverse spin-fields insertions $\tilde{S}_{a_1}^{\text{tr}}(\bar{z}_1), \dots, \tilde{S}_{a_s}^{\text{tr}}(\bar{z}_s)$ with fixed triality-rotated $SO(8)$ indices $a_1, \dots, a_s$. For $n \leq 3$ the 8-vectors $v_{a_1}, \dots, v_{a_s}$ span a vector space $V \subset \mathbb{R}^8$ of dimension $d \leq s \leq 2n \leq 6 < 8$, and there are $2(8-d) \geq 4$ supercharges

$$\tilde{Q}_{(w, \pm)} = w^b \oint \frac{d\bar{z}}{2\pi} e^{\pm \frac{i}{2} \tilde{\varphi} - \frac{1}{2} \tilde{\phi}} \tilde{S}_b^{\text{tr}}(\bar{z}), \quad \text{with } w \cdot v_{a_i} = 0 \text{ for } i = 1, \dots, s, \quad (10.97)$$

which anticommute with all inserted $\tilde{S}_{a_i}^{\text{tr}}$ while $\tilde{Q}_{(w,\pm)}^\dagger = \tilde{Q}_{(w,\mp)}$. The Hermitian supercharges $\tilde{Q}_{(w,+)} + \tilde{Q}_{(w,-)}$ commute with the operator $\mathcal{U}_{(a_1,\dots,a_s)}$ and, by the previous argument, we conclude that $\text{Str } \mathcal{U} = 0$ for $n \leq 3$ which is our non-renormalization theorem.

References

1. Verlinde, E., Verlinde, H.: Chiral bosonization, determinants and string partition functions. Nucl. Phys. B. **288**, 357–396 (1987)
2. Verlinde, E., Verlinde, H.: Multiloop calculations in covariant superstring theory. Phys. Lett. B. **192**, 95–102 (1987)
3. Lechtenfeld, O.: Superconformal ghost correlations on Riemann surfaces. Phys. Lett. B. **232**, 193–198 (1989)
4. Atick, J.J., Rabin, J.M., Sen, A.: An ambiguity in fermionic string perturbation theory. Nucl. Phys. B. **299**, 279–294 (1988)
5. Atick, J.J., Rabin, J.M., Sen, A.: Catoptric tadpoles. Nucl. Phys. B. **307**, 221–273 (1988)
6. Verlinde, E.P., Verlinde, H.L.: Lectures on string perturbation theory, Preprint IASSNS-HEP-88-52
7. Witten, E.: Notes on super-Riemann surfaces and their moduli. Pure Appl. Math. Quart. **15**(1), 57–211 (2019). [arXiv:1209.2459](#)
8. Witten, E.: Superstring perturbation theory revisited. [arXiv:1209.5461](#)
9. Witten, E.: Notes on supermanifolds and integration. Pure Appl. Math. Quart. **15**(1), 3–56 (2019). [arXiv:1209.2199](#)
10. Witten, E.: Superstring perturbation theory via super-Riemann surfaces: an overview. [arXiv:1304.2832](#)
11. Witten, E.: Notes on holomorphic string and superstring theory measures of low genus. [arXiv:1306.3621](#)
12. Berezin, F.A., Kirillov, A.A., Reidel, D. (eds.): Introduction to Superanalysis (1987)
13. Manin, Y.I.: Gauge Field Theory and Complex Geometry. Springer-Verlag (1988)
14. DeWitt, B.: Supermanifolds, 2nd edn. Cambridge University Press (1992)
15. Rogers, A.: Supermanifolds: Theory and Applications. World-Scientific (2007)
16. Baranov, M.A., Frolov, I.V., Shvarts, A.S.: Geometry of two-dimensional superconformal field theories. Theor. Math. Phys. **70**, 64–72 (1987)
17. Donagi, R., Witten, E.: Supermoduli space is not projected. Proc. Symp. Pure Math. **90**, 19–72 (2015). [arXiv:1304.7798](#)
18. Sen, A., Witten, E.: Filling the gaps with PCO's. JHEP. **1509**, 004 (2015). [arXiv:1504.00609](#)
19. La, H.S., Nelson, P.C.: Unambiguous fermionic string amplitudes. Phys. Rev. Lett. **63**, 24 (1989)
20. Seiberg, N.: Anomalous dimensions and mass renormalization in string theory. Phys. Lett. B. **187**, 56 (1987)
21. Sen, A.: Mass renormalizations and BRST anomaly in string theories. Nucl. Phys. B. **304**, 403–416 (1988)
22. Pius, R., Rudra, A., Sen, A.: Mass renormalization in string theory: general states. JHEP. **1407**, 062 (2014). [arXiv:1401.7014](#)
23. Sen, A.: Off-shell amplitudes in superstring theory. Fortsch. Phys. **63**, 149–188 (2015). [arXiv:1408.0571](#)
24. Sen, A.: One loop mass renormalization of unstable particles in superstring theory. JEPH. **1611**, 050 (2016). [arXiv:1607.06500](#)
25. Sen, A.: Unitarity of superstring field theory. JHEP. **1612**, 115 (2016). [arXiv:1607.06244](#)

26. Sen, A.: Wilsonian effective action of superstring theory. *JHEP.* **1701**, 108 (2017). [arXiv:1609.00459](#)
27. Sen, A.: Background independence of closed superstring field theory. [arXiv:1711.08468](#)
28. Lerche, W., Nilsson, B.E.W., Schellekens, A.N., Warner, N.P.: Anomaly cancelling terms from the elliptic genus. *Nucl. Phys. B.* **299**, 91–116 (1988)
29. Green, M.B., Schwarz, J.H.: Covariant description of superstrings. *Phys. Lett. B.* **136**, 367 (1984)
30. Dine, M., Seiberg, N.: Nonrenormalization theorems in superstring theory. *Phys. Rev. Lett.* **57**, 2625–2628 (1986)
31. Martinec, E.J.: Nonrenormalization theorems and fermionic string finiteness. *Phys. Lett. B.* **171**, 189 (1986)

Chapter 11

Calabi–Yau Compactifications



Abstract We study a large class of *exact* superstring vacua obtained by compactifying a 10d SUSY string down to 4d on Riemannian spaces of $SU(3)$ holonomy a.k.a. *Calabi–Yau* (CY) *3-folds*. These vacua are invariant under 4d Poincaré symmetry and one-quarter of the original supersymmetries. We discuss their deep geometry and physics, aiming to be didactical but also mathematically precise and self-contained. We review all geometry we need. One may study these vacua from three viewpoints:

- (a) in Algebro-Geometric terms, exploiting the description of the Calabi–Yau spaces as complex projective varieties. The main tool here is Griffiths’ theory of *variations of Hodge structures* (VHS) [1–3];
- (b) in terms of the Zamolodchikov geometry of the (2,2) SCFT living on the string world-sheet. Here the main tool is *tt* geometry* [4–6];
- (c) in terms of their low-energy effective 4d $\mathcal{N} = 2$ SUGRA. The main tool here is “special Kähler geometry” [7–12].

These three approaches, despite the difference of languages, perspectives, and tools, turn out to be essentially equivalent [13].

Calabi–Yau compactifications were introduced in [14–16].

11.1 Geometric Background

We review the required background from geometry. Experts may jump to Sect. 11.2.

11.1.1 Mini-Review of Differential Geometry (DG)

In this section, M is a smooth manifold of dimension n and $TM \rightarrow M$ its tangent bundle [17, 18]. We write $LM \rightarrow M$ for the bundle of linear frames on TM whose local sections are n -tuples of pointwise linearly independent vector fields. $LM \rightarrow M$ is a smooth $GL(n, \mathbb{R})$ -principal bundle [17, 18].

G-Structures on M Let $G \subset GL(n, \mathbb{R})$ be a *closed* Lie subgroup.

Definition 11.1 A G -structure on M is a smooth G -principal sub-bundle $P \subset LM$. The G -structure P is *integrable* iff every point of M has a neighborhood with local coordinates $x^1, \dots, x^n$ such that $(\partial/\partial x^1, \dots, \partial/\partial x^n)$ is a local section of P .

It follows from the definition that specifying a G -structure is the same as giving a section $s: M \rightarrow X_G$ where the bundle X_G has typical fiber $GL(n, \mathbb{R})/G$ [17].

The embedding $G \hookrightarrow GL(n, \mathbb{R})$ defines a representation ρ of G in $\mathbb{R}^n$. Let

$$\rho(P) \equiv P \times_{\rho} \mathbb{R}^n \stackrel{\text{def}}{=} (P \times \mathbb{R}^n)/[(p, v) \sim (pg^{-1}, \rho(g)v), \forall g \in G] \quad (11.1)$$

be the *associated vector bundle* [19]. P defines an isomorphism

$$e: TM \rightarrow \rho(P). \quad (11.2)$$

e is a section of the vector bundle $T^*M \otimes \rho(P)$, which we call the G -vielbein (the term *soldering form* is also used). In a local trivialization of $\rho(P)$, e is written as a n -tuple of 1-forms $e^a = e^a_{\mu} dx^{\mu}$ ($a = 1, \dots, n$). The *structure group* G acts on the index a via the matrices in the representation ρ .

Let $T \in (\otimes^k \mathbb{R}^n) \otimes (\otimes^{\ell} \mathbb{R}^n)^{\vee}$ be a tensor of type¹ (k, ℓ) in $\mathbb{R}^n$ invariant for G acting in the representation $\rho^{\otimes k} \otimes (\rho^{\vee})^{\otimes \ell}$. T defines a *tensor field* $\mathcal{T}$ on M in the following way: at each point $x \in M$ we choose a frame $u_x \in LM_x$ which yields a linear isomorphism $e_x: \mathbb{R}^n \rightarrow TM_x$, which extends to an isomorphism of tensor algebras. We set $\mathcal{T}_x = e_x(T)$; since T is G -invariant $\mathcal{T}_x$ does not depend on the choice of $e_x \in P_x$. In the jargon of GR one says that the vielbein $(e_x)_{\mu}^a \equiv e(x)_{\mu}^a$ transforms the “flat” indices a_s, b_t of T into “curved” indices μ_s, ν_t of $\mathcal{T}$

$$\begin{aligned} \mathcal{T}(x)_{\mu_1 \dots \mu_k}^{\nu_1 \dots \nu_{\ell}} &= \\ &= e(x)_{\mu_1}^{a_1} \dots e(x)_{\mu_k}^{a_k} e(x)^{\nu_1}_{b_1} \dots e(x)^{\nu_{\ell}}_{b_{\ell}} T^{b_1 b_2 \dots b_k}_{a_1 a_2 \dots a_{\ell}}. \end{aligned} \quad (11.3)$$

A *connection on a G -structure* is a connection on the principal bundle P . The representation ρ then induces a connection on the bundle $\rho(P) \simeq TM$ which locally (and in a chosen gauge) is written as a 1-form ω^a_b with coefficients in $\mathfrak{g} = \mathfrak{Lie}(G)$.

The *torsion* T^a and *curvature* R^a_b of the G -structure with connection ω^a_b are

$$T^a \stackrel{\text{def}}{=} \nabla e^a \equiv de^a + \omega^a_b \wedge e^b, \quad R^a_b \stackrel{\text{def}}{=} d\omega^a_b + \omega^a_c \wedge \omega^c_b. \quad (11.4)$$

They satisfy the Bianchi identities

$$\nabla T^a \equiv dT^a + \omega^a_b \wedge T^b = R^a_b \wedge e^b, \quad (11.5)$$

$$\nabla R^a_b \equiv dR^a_b + \omega^a_c \wedge R^c_b - R^a_c \wedge \omega^c_b = 0. \quad (11.6)$$

¹ That is, a tensor in $\mathbb{R}^n$ with the index structure $T^{b_1 b_2 \dots b_k}_{a_1 a_2 \dots a_{\ell}}$ with $a_s, b_t = 1, 2, \dots, n$.

We state two facts without proof (they may be found e.g. in [17, 19]):

Proposition 11.1 P integrable $\Rightarrow P$ admits a torsion-free connection.

Proposition 11.2 P integrable $\Leftrightarrow$ locally we can find coordinates such that all tensor fields $\mathcal{T}$ associated to G -invariant tensors T in $\mathbb{R}^n$ have constant components.

Definition 11.2 By a *torsionless G -structure* we mean a pair (θ, ω) with zero torsion $\nabla\theta^a = de^a + \omega^a_b \wedge e^b = 0$. A torsionless G -structure will be called *strict* iff the holonomy group² of its curvature R^a_b is (isogeneous) to the full group G and not to a proper subgroup, i.e. if the 2-form R^a_b takes values in $\mathfrak{g} \equiv \mathfrak{Lie}(G)$ and not in a proper Lie subalgebra. The G -structure with connection is called *locally symmetric* if $\nabla_i R_{jk}^a = 0$, i.e. if the curvature tensor is *parallel* ($\equiv$ covariantly constant).

Examples of G -Structure We give some very basic example of G -structure:

- (a) $G = GL(n, \mathbb{R})^+$, the group of matrices of positive determinant. A $GL(n, \mathbb{R})^+$ -structure is an *orientation of M* and it exists iff M is orientable. When it exists, a $GL(n, \mathbb{R})^+$ -structure is automatically integrable;
- (b) $G = SL(n, \mathbb{R})$. A $SL(n, \mathbb{R})$ -structure defines a *volume form* on M . It exists iff M is orientable and then is integrable;
- (c) $G = GL(n/2, \mathbb{C})$, n =even: an *almost complex structure* on M . When the $GL(n/2, \mathbb{C})$ -structure is integrable we say that M is a *complex manifold*;
- (d) $G = Sp(n, \mathbb{R})$, n =even: an *almost symplectic structure*. Invariant tensor a 2-form Ω . M admits an almost symplectic structure iff it has an almost complex structure. When the structure is integrable, M is a *symplectic manifold*;
- (e) $G = O(n)$. The invariant tensor is a positive metric $g_{\mu\nu} = e^a_\mu e^b_\nu \delta_{ab}$, so an $O(n)$ -structure is a *Riemannian metric*. Integrable iff the metric is flat;
- (f) G =(any compact subgroup). Since $O(n)$ is the maximal compact subgroup of $GL(n, \mathbb{R})$, $G \subseteq O(n)$ and hence the G -structure defines (in particular) a Riemannian metric: we get a *special Riemannian geometry*;
- (g) $G = U(n/2)$, n even: an *almost Hermitian structure*. Since

$$U(n/2) \subset GL(n/2, \mathbb{C}), \quad U(n/2) \subset SO(n), \quad U(n/2) \subset Sp(n, \mathbb{R}) \quad (11.7)$$

a $U(n/2)$ -structure defines an almost complex structure I , a Riemannian metric g , and an almost symplectic structure Ω .

Definition 11.3 A *Kähler structure* is a torsionless $U(n/2)$ -structure.

Fundamental Theorems We list without proof the fundamental theorems of DG:

Theorem 11.1 (Newlander–Nirenberg [19, 20]) A $GL(n/2, \mathbb{C})$ -structure is integrable iff it admits a torsion-free connection.

² Properly speaking, this is the restricted holonomy groups, i.e. its connected component.

Theorem 11.2 (Darboux [19]) **(1)** *An almost symplectic structure is a symplectic structure iff the invariant 2-form Ω is closed.* **(2)** *An almost symplectic structure is a symplectic structure iff it admits a torsionless connection.*

Theorem 11.3 (The fundamental theorem of DG [21]) *On a $SO(n)$ -structure there is a **unique** connection of given torsion T^a . In particular there is a **unique** torsionless metric connection, the Levi-Civita one.*

Corollary 11.1 *A Kähler structure, i.e. a torsionless $U(n/2)$ -structure, is a special torsionless $SO(n)$ -structure, so its connection is the Levi-Civita one, while the 2-form Ω is closed by the Darboux theorem, and the underlying complex structure I is integrable by the Newlander–Nirenberg, so M is a complex manifold, endowed with a Hermitian metric whose associated 2-form (the Kähler form) is closed. Moreover the Levi-Civita connection is the Chern one³ on the holomorphic tangent bundle.*

The statements in the **Corollary** are often taken as the definition of Kähler manifold.

Riemannian Holonomy Groups and All That For a nice account see [23].⁴

Let M be a connected Riemannian manifold, and fix a base point $x_0 \in M$. Levi-Civita parallel transport of vectors along a loop ℓ , based at x_0 , defines an element $W_\ell \in O(T_{x_0}M) \simeq O(n)$. The *holonomy group* $\mathfrak{Hol}$ is the closure in $O(n)$ of the group generated by W_ℓ for all ℓ . The Riemannian structure defines a torsionless $\mathfrak{Hol}$ -structure on M , and $R^a{}_b$ is a 2-form valued in the holonomy Lie algebra $\mathfrak{hol}$.

Theorem 11.4 (de Rham) *M carries a **strict** torsionless G -structure with G **compact**. Assume M to be **simply-connected** and **complete** for the associated Riemannian metric. If the representation ρ is **reducible**, $\rho = \rho_1 \oplus \rho_2$, then $G \equiv G_1 \times G_2$ with ρ_a a representation of G_a ($a = 1, 2$), and $M = M_1 \times M_2$ (Riemannian product), where M_a ($a = 1, 2$) has dimension $\dim \rho_a$ and a strict torsionless G_a -structure.*

Thus, when G is compact and M complete, going to the universal cover $\tilde{M}$ if necessary, we may assume the holonomy representation ρ **irreducible** without loss. In this case, we say that the Riemannian manifold M is **irreducible**.

Theorem 11.5 (Berger) *G compact. M carries a torsionless, irreducible, and **non** locally symmetric G -structure. Then the pair (G, ρ) is one of the following*

$\dim M$	G	ρ	name of G -manifold
n	$SO(n)$	n	Riemannian
$2m$	$U(m)$	m	Kähler (K)
$2m$	$SU(m)$	m	Calabi–Yau (CY)
$4m$	$Sp(m)$	$2m$	hyperKähler (HK)
$4m$	$Sp(1) \times Sp(m)$	$(2, m)$	Quaternionic Kähler (QK)
7	G_2	7	G_2 -manifold
8	$Spin(7)$	8	$Spin(7)$ -manifold

³ Recall that the Chern connection [22] is the *unique* connection on a holomorphic vector bundle with a Hermitian fiber metric which is both holomorphic and metric.

⁴ For more details see [24, 25]. For a readable summary for physicists see Chap. 3 of [7].

The groups G in the above *Berger list* do not preserve any *symmetric* tensor except the metric and its symmetrized tensor products. The group embeddings

$$\{1\} \subset Sp(m) \subset SU(2m) \subset U(m) \quad (11.8)$$

imply the embedding of geometries (as special cases)

$$(\text{flat} \Rightarrow \text{HK} \Rightarrow \text{CY} \Rightarrow \text{K}), \quad (\text{HK} \Rightarrow \text{QK}). \quad (11.9)$$

The Holonomy Principle Suppose the Riemannian manifold M has a parallel (i.e. covariantly-constant⁵) tensor T^A (A a multi-index standing for k covariant and ℓ contravariant indices): $\nabla_i T^A = 0$. Then M is endowed with a torsionless G -structure whose group G is contained in the subgroup of $SO(n)$ leaving invariant T^A . Indeed

$$0 = [\nabla_i, \nabla_j]T^A = R_{ij}{}^A{}_B T^B, \quad (11.10)$$

so T^A is invariant under the holonomy Lie algebra $\mathfrak{hol}$.

Example: Irreducible with a parallel 2-form $\Rightarrow$ Kähler

M irreducible, non-symmetric with a parallel 2-form $\Omega = \Omega_{ij}dx^i \wedge dx^j$. The tensor $\Omega_{ik}g^{kl}\Omega_{lj}$ is symmetric and parallel, so a multiple of the metric,

$$\Omega_{ik}g^{kl}\Omega_{lj} = \lambda g_{ij} \quad \text{with} \quad \lambda < 0. \quad (11.11)$$

We normalize Ω so that $\lambda = -1$. $I^i{}_j \equiv g^{ik}\Omega_{kj}$ is an almost complex structure, $I^2 = -\text{Id}$, integrable by the Newlander–Nirenberg theorem, while the 2-form Ω is closed (being parallel). Thus: *An irreducible Riemannian manifold M has a parallel 2-form if and only if it is Kähler.* An irreducible symmetric manifold with a parallel 2-form is both symmetric and Kähler, i.e. a Cartan Hermitian space [26].

More generally, a non-symmetric, irreducible Riemannian manifold has a *special* holonomy group (i.e. *strict* $G \subsetneq SO(n)$) iff it has the following parallel forms

G	name	parallel forms
$U(m)$	Kähler	ω^k with $k = 0, \dots, m$; ω = Kähler 2-form
$SU(m)$	Calabi–Yau	$\omega^k, \epsilon, \bar{\epsilon}$; ϵ = holomorphic $(m, 0)$ form
$Sp(m)$	hyperKähler	polynomials in the 3 Kähler forms ω_a
$Sp(1) \times Sp(m)$	quaternionic Kähler	$\Theta^k, k = 0, \dots, m$, with Θ a 4-form
G_2		a 3-form ϕ and its dual 4-form $*\phi$
$Spin(7)$		a self-dual 4-form ϕ

as one easily checks by elementary group theory. The holonomy principle extends to *parallel spinors*; indeed⁶

⁵ Here and below *parallel* stands for covariantly constant with respect to the Levi-Civita connection.

⁶ Here Γ^i are Dirac matrices acting on a spin bundle S on the manifold M and $\Gamma^{ij} \equiv \frac{1}{2}[\Gamma^i, \Gamma^j]$. The generators of $\mathfrak{spin}(\dim M)$ acting on a spinor $\psi \in C^\infty(M, S)$ are the Dirac matrices $\frac{1}{2}\Gamma^{ij}$.

$$\nabla_i \psi = 0 \Rightarrow 0 = [\nabla_i, \nabla_j] \psi = \frac{1}{4} R_{ijkl} \Gamma^{kl} \psi \quad (11.12)$$

so that $\psi \neq 0$ implies that ψ is G -invariant and then G must be special.

Theorem 11.6 *M has a non-zero parallel spinor $\Rightarrow M$ is Ricci-flat.*

Proof From Eq. (11.12) and the 1st Bianchi identity

$$\begin{aligned} 0 &= R_{ijkl} \Gamma^j \Gamma^{kl} \psi = R_{ijkl} (\Gamma^{jkl} - \delta^{jk} \Gamma^l + \delta^{jl} \Gamma^k) \psi = 2 R_{ij} \Gamma^j \psi \\ \Rightarrow 0 &= g^{ij} R_{ik} R_{jl} \Gamma^k \Gamma^l \psi = g^{ij} g^{kl} R_{ik} R_{jl} \psi \Rightarrow R_{ij} = 0. \end{aligned} \quad (11.13)$$

Note that $R_{ij} R^{ij} = 0$ implies $R_{ij} = 0$ in positive signature but not in the Lorentzian one: thus we can have non-Ricci-flat Lorentzian metrics with parallel spinors. $\square$

If M has a parallel spinor and is locally symmetric, then M is flat since a Ricci-flat locally symmetric space is flat [23]. Otherwise we may assume M irreducible non-locally symmetric without loss. Then G is either trivial or a group in the Berger list.

By the holonomy principle, the number of parallel spinors is the number of trivial representations in the decomposition of the spinor representation S of $Spin(n)$ under the subgroup G . If n is even, we consider separately the two chiral representations $S_{\pm}$. E.g. for $G_2 \subset Spin(7)$ the spinor representation decomposes as $\mathbf{8} = \mathbf{7} \oplus \mathbf{1}$ and we have *one parallel spinor*. For $Spin(7) \subset Spin(8)$ we have $\mathbf{8}_s = \mathbf{8}$ and $\mathbf{8}_c = \mathbf{7} \oplus \mathbf{1}$, so we have *a single parallel spinor of negative chirality*.

Parallel Spinors for M Kähler We focus on $G = U(m)$, i.e. M Kähler. Using θ we identify the γ -matrices on M with the ones on $\mathbb{R}^{2m} \simeq \mathbb{C}^m$ which have a $U(m)$ -invariant splitting into $(1, 0) \oplus (0, 1)$ types, $\Gamma^j, \Gamma^{\bar{k}}$. The Clifford algebra reads

$$\Gamma^j \Gamma^{\bar{k}} + \Gamma^{\bar{k}} \Gamma^j = 2 \delta^{j\bar{k}}, \quad \Gamma^i \Gamma^j + \Gamma^j \Gamma^i = \Gamma^{\bar{k}} \Gamma^{\bar{l}} + \Gamma^{\bar{l}} \Gamma^{\bar{k}} = 0. \quad (11.14)$$

The spinorial representation is constructed by picking a Clifford vacuum ψ_0 satisfying $\Gamma^{\bar{k}} \psi_0 = 0$ for all $\bar{k}$. The chiral spinor spaces S_+, S_- are spanned by the vectors

$$\Gamma^{i_1 i_2 \dots i_s} \psi_0 \quad (11.15)$$

with s even or odd, respectively. By “PCT symmetry”⁷ the $U(1) \subset U(m)$ charge of (11.15) is $(s - m/2)$. There are only 2 vectors invariant under the subgroup $SU(m)$

$$\psi_0 \quad \text{and} \quad \Gamma^{12 \dots m} \psi_0. \quad (11.16)$$

Their $U(1)$ charges are $\pm m/2$, so no element of $S_{\pm}$ is invariant, and a strict Kähler manifold has no parallel spinor. The argument shows that for $G = SU(m)$, i.e. M a strict CY, we have 2 parallel spinors ψ_0 and $\Gamma^{1 \dots m} \psi_0$ of the same chirality for m even and opposite chirality for m odd.

⁷ Indeed the spectrum of $U(1)$ charges q of the states (11.15) should be symmetric under $q \rightarrow -q$, and this fixes the charge of ψ_0 to be $-m/2$.

When $G = Sp(m) \subset SU(2m)$ we have a $(2, 0)$ parallel form, i.e. the holomorphic symplectic form Ω_{ij} . Let $\omega \equiv \Omega_{ij}\Gamma^i\Gamma^j$. By the previous analysis, we have the following list of $Sp(m)$ singlets

$$\psi_0, \quad \omega\psi_0, \quad \omega^2\psi_0, \quad \dots, \quad \omega^m\psi_0 \quad (11.17)$$

i.e. a strict HK of $\mathbb{R}$ -dimension $4m$ has $m + 1$ parallel spinors, all of the same chirality.

When $G = Sp(1) \times Sp(m)$ the parallel forms should be, in particular, $Sp(m)$ invariant. But the list (11.17) of $Sp(m)$ -invariant spinors transforms in the irreducible $\mathbf{m} + \mathbf{1}$ representation of $Sp(1)$, so no element is invariant under the full $Sp(1) \times Sp(m)$. A strict QK has *no* parallel spinor.⁸

Corollary 11.2 *CY manifolds, HK manifolds, G_2 -manifolds, and $Spin(7)$ -manifolds are Ricci-flat, $R_{ij} = 0$, i.e. Euclidean solutions to the vacuum Einstein equations. The number of $n_{\pm}$ of parallel spinors of given chirality is⁹*

holonomy group	$SU(2m+1)$	$SU(2m)$	$Sp(m)$	G_2	$Spin(7)$
(n_+, n_-)	$(1, 1)$	$(2, 0)$	$(m+1, 0)$	I	$(0, 1)$

Corollary 11.3 *Let M be a compact Riemannian manifold with non-zero parallel spinors. Then, up to a **finite** cover, M has the form*

$$T^n \times M_1 \times M_2 \times \dots \times M_s, \quad (11.18)$$

where the M_i are simply-connected, compact, irreducible, strict, torsionless G_i -manifolds with $G_i \in \{SU(m_i), Sp(k_i), G_2, Spin(7)\}$, and T^n is a flat torus. Moreover

$$\#(\text{parallel spinors in } \mathbb{R}^{d-1,1} \times M) = 2^{[(d+n)/2] + \#CY} \cdot \prod_{HK} (k_i + 1) \quad (11.19)$$

where $\#CY$ is the number of strict CY factors and the product is over the HK factors.

Corollary 11.4 *M a non-necessarily strict CY m -fold. Let $\Omega^{*,0} \equiv \bigoplus_{p=0}^m \Omega^{p,0}$ be the graded ring of smooth $(p, 0)$ -forms, $0 \leq p \leq m$, and S the space of smooth spinor fields (sections of the spin bundle). Let $\psi_0 \in S$ be a parallel spinor with $\Gamma^i\psi_0 = 0$ (unique up to normalization). Then*

$$\Omega^{*,0} \simeq S \quad \text{isomorphism of } \Omega^{*,0}\text{-modules} \quad (11.20)$$

$$\phi_{i_1 i_2 \dots i_p} dz^{i_1} \wedge dz^{i_2} \wedge \dots \wedge dz^{i_p} \longmapsto \phi_{i_1 i_2 \dots i_p} \Gamma^{i_1 i_2 \dots i_p} \psi_0. \quad (11.21)$$

⁸ This is clear from the fact that a strict QK manifold is Einstein with a non-zero cosmological constant—hence has no parallel spinor by **Theorem 11.6**. A Ricci-flat QK manifold is hyperKähler.

⁹ A G_2 -manifold has odd dimension (7) and there is no notion of chirality for parallel spinors.

Under this isomorphism the curved-space Dirac operator $\mathcal{D}$ is mapped into the Dolbeault–Kähler–Dirac operator $\partial + \partial^\dagger$. If M is compact, all solutions to the Dirac equation $\mathcal{D}\psi = 0$ are parallel spinors; thus all holomorphic $(p, 0)$ -forms are parallel.

Exercise 11.1 Prove last statement of **Corollary**.

In particular if X is strict CY we have the correspondence

$$\{1, \epsilon\} \leftrightarrow \{\psi_0, \Gamma^1 \cdots \Gamma^m \psi_0\} \quad (11.22)$$

between parallel forms and spinors. The isomorphism in **Corollary 11.4** plays a crucial role in 2d SUSY QFT where is called the *spectral-flow isomorphism* [27].

11.1.2 Complex and Kähler Manifolds

Since $SU(m) \subset U(m) \subset GL(m, \mathbb{C})$, a Riemannian manifold M with holonomy $SU(m)$ —i.e. with a torsionless $SU(m)$ -structure—is in particular Kähler, hence complex. A manifold M is complex iff it is equipped with an integrable $GL(m, \mathbb{C})$ -structure, that is, iff there is an open cover $M = \cup_\alpha U_\alpha$ and local complex coordinates $z_\alpha^1, \dots, z_\alpha^m$ in each patch U_α , with holomorphic Jacobians $\partial z_\beta^i / \partial z_\alpha^j$ in the overlaps $U_\alpha \cap U_\beta$. The complexified tangent bundle splits into types¹⁰ $(1, 0)$ and $(0, 1)$:

$$T_{\mathbb{R}}M \otimes_{\mathbb{R}} \mathbb{C} = T^{1,0}M \oplus T^{0,1}M, \quad T^{0,1}M \equiv \overline{T^{1,0}M}. \quad (11.23)$$

The smooth sections of $T^{1,0}M$ are complex vector fields of the local form $v^i \partial_{z^i}$ for C^∞ coefficients v^i . $T^{1,0}M \rightarrow M$ is a *holomorphic vector bundle* [19, 20, 22, 28, 29], i.e. its transition functions $\partial z_\beta^i / \partial z_\alpha^j$ are holomorphic in $U_\alpha \cap U_\beta$. Dually, the complexified cotangent bundle splits into types $(1, 0)$ and $(0, 1)$

$$T_{\mathbb{R}}^*M \otimes \mathbb{C} = \Lambda^{1,0} \oplus \Lambda^{0,1}, \quad \Lambda^{0,1} = \overline{\Lambda^{1,0}}, \quad (11.24)$$

and the bundle of complex k -forms splits in sub-bundles of *definite type* (p, q)

$$\Omega^k \equiv \wedge^k T^*M \otimes \mathbb{C} = \bigoplus_{p+q=k} \Lambda^{p,q}, \quad \Lambda^{q,p} = \overline{\Lambda^{p,q}}. \quad (11.25)$$

The local sections of $\Lambda^{p,q}$ are smooth $(p+q)$ -forms of the form

$$\phi_{i_1 \dots i_p j_1 \dots j_q} dz^{i_1} \wedge \cdots \wedge dz^{i_p} \wedge d\bar{z}^{j_1} \wedge \cdots \wedge d\bar{z}^{j_q}. \quad (11.26)$$

The bundle $\Lambda^{p,0} \rightarrow M$ is holomorphic for all p . $\mathcal{K} \equiv \Lambda^{m,0}$ is a holomorphic line bundle called the *canonical line bundle of M* .

¹⁰ The overbar stands for complex conjugation.

We assume that our complex manifold M is equipped with a Hermitian metric.¹¹ Decomposing into type the exterior derivative d we have

$$d = \partial + \bar{\partial}, \quad \delta \equiv d^\dagger = \partial^\dagger + \bar{\partial}^\dagger, \quad (11.27)$$

$$\begin{cases} \partial: \Lambda^{p,q} \rightarrow \Lambda^{p+1,q}, & \bar{\partial}: \Lambda^{p,q} \rightarrow \Lambda^{p,q+1} \\ \partial^\dagger: \Lambda^{p,q} \rightarrow \Lambda^{p-1,q}, & \bar{\partial}^\dagger: \Lambda^{p,q} \rightarrow \Lambda^{p,q-1} \end{cases} \quad (11.28)$$

where $\dagger$ is the formal Hodge adjoint, $a^\dagger \equiv *^{-1}a*$, with $*$: $\Lambda^{p,q} \rightarrow \Lambda^{n-p,n-q}$ the Hodge dual defined by the Hermitian metric [20]. The operator $\bar{\partial}$ is called the *Dolbeault differential* (or operator). It satisfies

$$d^2 = 0 \Rightarrow \partial^2 = \bar{\partial}^2 = \partial\bar{\partial} + \bar{\partial}\partial = 0. \quad (11.29)$$

The Dolbeault cohomology groups are defined as [19, 20, 22, 28, 29]

$$H^{p,q}(M) \stackrel{\text{def}}{=} \ker(\bar{\partial}: \Lambda^{p,q} \rightarrow \Lambda^{p,q+1}) / \text{im}(\bar{\partial}: \Lambda^{p,q-1} \rightarrow \Lambda^{p,q}). \quad (11.30)$$

Kähler Manifolds (Torsionless $U(m)$ -Structures)

We need some basic facts about Kähler manifolds. To make the story ultra-short, we use the physical characterization of Kähler manifolds as target spaces of σ -models with 4-supercharge SUSY (cf. Sect. 8.1). We write $\omega \equiv ig_{i\bar{j}}dz^i \wedge d\bar{z}^{\bar{j}}$ for the Kähler form. One has $d\omega = 0$ (**Corollary 11.1**).

Proposition 11.3 (see e.g. [20, 22, 28, 30]) *On a Kähler manifold M the following Kähler-Hodge identities hold (Δ is the Hodge Laplacian):*

$$\partial^2 = \bar{\partial}^2 = \partial\bar{\partial} + \bar{\partial}\partial = \partial^{\dagger 2} = \bar{\partial}^{\dagger 2} = \partial^\dagger\bar{\partial}^\dagger + \bar{\partial}^\dagger\partial^\dagger = 0 \quad (11.31)$$

$$\partial\bar{\partial}^\dagger + \bar{\partial}^\dagger\partial = \partial^\dagger\bar{\partial} + \bar{\partial}\partial^\dagger = 0 \quad (11.32)$$

$$\partial\partial^\dagger + \partial^\dagger\partial = \bar{\partial}\bar{\partial}^\dagger + \bar{\partial}^\dagger\bar{\partial} = \frac{1}{2}(dd^\dagger + d^\dagger d) \equiv \frac{1}{2}\Delta. \quad (11.33)$$

Proof Consider the 4-SUSY 1d dimensional σ -model with target M . It is a QM system: its Schrödinger picture Hilbert space is isomorphic to the space of differential forms on M (with L^2 -coefficients) endowed with the Hodge Hermitian product [31, 32]. The Hamiltonian is a geometric intrinsic 2nd-order differential operator,

$$H = \frac{1}{2}p^i p^i + \dots \quad (11.34)$$

and the supercharges are four 1st-order differential operators. By functoriality and uniqueness, the Schrödinger representation of the 4 supercharges should be given by the 4 operators $\partial, \bar{\partial}, \partial^\dagger, \bar{\partial}^\dagger$.

¹¹ That is, a $U(n)$ -structure, possibly with a torsion, whose underlying $GL(n, \mathbb{C})$ -structure admits a torsionless connection.

The Kähler-Hodge identities (11.31)–(11.33) then follow from the 1d 4-SUSY algebra [31, 32]

$$\{Q_a, Q_b^\dagger\} = \delta_{ab} H, \quad H \equiv \frac{1}{2} \Delta. \quad (11.35)$$

□

A form ϕ is *harmonic* iff $\Delta\phi = 0$; $\mathbb{H}^k(M)$ is the vector space of harmonic k -forms on M . In a compact manifold all de Rham cohomology class has a **unique** harmonic representative, so $H_{\text{dR}}^k(M) \simeq \mathbb{H}^k(M)$ [28, 30]. Likewise, if M is compact and Hermitian, all Dolbeault cohomology class has a unique $\bar{\partial}$ -harmonic representative $\psi \in \mathbb{H}^{p,q}(M)$ such that $(\bar{\partial}\bar{\partial}^\dagger + \bar{\partial}^\dagger\bar{\partial})\psi = 0$. From Eq. (11.33) we get the following:

Corollary 11.5 *In a Kähler manifold M , the harmonic forms may be chosen to have definite type (p, q) . If M is compact, we have the Hodge decomposition*

$$H^k(M, \mathbb{C}) \equiv H_{\text{dR}}^k(M) \otimes \mathbb{C} = \bigoplus_{p+q=k} H^{p,q}(M), \quad H^{q,p}(M) = \overline{H^{p,q}(M)}. \quad (11.36)$$

The *Hodge numbers* of M are defined as

$$h^{p,q} \stackrel{\text{def}}{=} \dim_{\mathbb{C}} H^{p,q}(M) \equiv \dim_{\mathbb{C}} \mathbb{H}^{p,q}(M), \quad (11.37)$$

and have the properties (see e.g. [20, 22, 28, 30]):

$$\begin{array}{ll} \text{conjugation} & \text{Poincaré duality} \\ h^{p,q} = h^{q,p}, & h^{m-q, m-p} = h^{p,q}, \end{array} \quad (11.38)$$

$$\begin{array}{ll} \text{Bettin numbers} & \text{Lefschetz's } sl(2) \\ \sum_p h^{p, k-p} = b_k & h^{p-1, q-1} \leq h^{p,q} \text{ for } p+q \leq m. \end{array} \quad (11.39)$$

The $\{h^{p,q}\}$ are usually written as a diamond-shaped table called *Hodge diamond*.

Corollary 11.6 *In a compact Kähler manifold the odd Betti numbers are even and the even ones are positive ($\omega^k \in H^{k,k}(M)$ is parallel, hence harmonic).*

Example: A complex manifold which does not admit any Kähler metric

Take $\mathbb{C}^2 \setminus \{0, 0\} \simeq_{\text{smooth}} S^3 \times \mathbb{R}$ and quotient it by the equivalence $(z^1, z^2) \sim \lambda(z^1, z^2)$ with $\lambda > 1$. The quotient is a compact complex manifold (a special case of *Hopf surface* [33]) topologically equivalent to $S^3 \times \mathbb{R}/\mathbb{Z} \equiv S^3 \times S^1$ with Betti numbers $b_2 = 0$, $b_1 = 1$ and $b_3 = 1$. The Hopf surface cannot admit any Kähler metric by the **Corollary**.

From the Kähler-Hodge identities and the uniqueness of the Chern connection, it is obvious that a Hermitian metric on a complex manifold

$$ds^2 = g(z, \bar{z})_{j\bar{k}} dz^j d\bar{z}^k \quad (11.40)$$

is Kähler if and only if its Levi-Civita connection is the Chern connection on $T^{1,0}M$; hence the only non-zero Christoffel symbols are

$$\Gamma_{jk}^i \equiv g^{i\bar{l}} \partial_i g_{\bar{l}k} \quad \text{and its conjugate} \quad \Gamma_{\bar{j}\bar{k}}^{\bar{i}}. \quad (11.41)$$

The only non-zero components of the Riemann tensor are

$$R_{i\bar{j}k\bar{l}} = R_{i\bar{l}k\bar{j}} = R_{k\bar{j}i\bar{l}} = g_{\bar{l}m} \partial_{\bar{j}} \Gamma_{ik}^m, \quad (11.42)$$

and those obtained from these by the usual symmetries of the Riemann tensor. The Ricci tensor is of type (1, 1)

$$R_{i\bar{k}} = g^{j\bar{l}} R_{j\bar{k}l\bar{i}} = g^{j\bar{l}} R_{\bar{k}ij\bar{l}}. \quad (11.43)$$

Thus the *Ricci form* $\mathbf{Ric} \equiv i R_{i\bar{k}} dz^i \wedge d\bar{z}^k$ is the $\mathfrak{u}(1)$ part of the $\mathfrak{u}(n) \equiv \mathfrak{u}(1) \oplus \mathfrak{su}(n)$ curvature. Therefore for a Kähler manifold

Ricci curvature vanishes $\Leftrightarrow$ holonomy reduces to $SU(n) \Leftrightarrow$ metric is CY

that is, Calabi–Yau means Kähler and Ricci-flat. We have the explicit formula

$$R_{i\bar{j}} = -\partial_{\bar{j}} \Gamma_{ik}^k = -\partial_{\bar{j}} g^{k\bar{l}} \partial_i g_{\bar{l}k} \equiv -\partial_{\bar{j}} \partial_i \log \det g_{\bar{l}k}. \quad (11.44)$$

By definition, $\det g$ is the Hermitian metric on the anti-canonical bundle $\mathcal{K}^{-1} \equiv \wedge^m T^{1,0}M$, so the Ricci form

$$\mathbf{Ric} = -i(\partial_i \partial_{\bar{j}} \log \det g) dz^i \wedge d\bar{z}^j \quad (11.45)$$

is minus the Chern curvature of the canonical bundle, i.e. $\mathbf{Ric}/2\pi$ represents the *first Chern class* $c_1(M)$ of the holomorphic tangent bundle $T^{1,0}M$.

11.1.3 Calabi–Yau Manifolds (CY)

In this section X is a compact Calabi–Yau of complex dimension m . It has at least two parallel spinors, and exactly two if M is a *strict* CY. X is in particular Ricci-flat, so an Euclidean solution to the vacuum Einstein equations.

Caveat. In the literature, there are a dozen inequivalent definitions of “Calabi–Yau manifold” [34]. Beware which one is meant in each specific argument.

Fact 11.1 A compact Kähler manifold M is CY $\Rightarrow$ its canonical bundle $\mathcal{K}$ is holomorphically trivial, $\mathcal{K} \simeq \mathcal{O}$, hence $c_1(M) = 0$.

Proof By the relation between holonomy and parallel forms, a complex manifold is CY iff it has a parallel (hence nowhere vanishing) section ϵ of $\Lambda^{m,0} \equiv \mathcal{K}$. Being parallel, ϵ is $\bar{\partial}$ -closed, so holomorphic. A global holomorphic section without zeros yields an isomorphism with the trivial bundle $\mathcal{O} \rightarrow \mathcal{K}$, $\phi \mapsto \phi \epsilon$. $\square$

Fact 11.2 X compact CY with $H^1(X) \neq 0$. Then (up to a finite cover) $X = T^2 \times X'$ with T^2 a 2-torus and X' a compact Calabi–Yau of complex dimension $\dim_{\mathbb{C}} X - 1$.

Proof X is in particular Kähler, so $H^1(X) \neq 0 \Rightarrow h^{1,0} \equiv \frac{1}{2}b_1 \geq 1$. By **Corollary 11.4** the $h^{1,0}$ holomorphic (1,0)-forms are parallel $\Rightarrow$ the holonomy representation ρ contains $h^{1,0} \geq 1$ copies of the trivial representation. Now apply **Theorem 11.4**. Alternatively: use **Corollary 11.3**. $\square$

Corollary 11.7 $H^1(X) = 0$ and $1 < \dim_{\mathbb{C}} X \leq 3 \Rightarrow X$ is a strict CY.

From **Corollary 11.4**, for a strict CY of complex dimension m ,

$$h^{p,0} = \begin{cases} 1 & p = 0, m \\ 0 & \text{otherwise.} \end{cases} \quad (11.46)$$

In mathematics, it is convenient to define a Calabi–Yau m -fold (m -CY) as a

(*) compact complex m -fold with $\mathcal{K}$ trivial and $h^{1,0} = 0$ which admits a Kähler metric

This definition has the merit of focusing on the underlying complex manifold, without reference to a specific metric of $SU(m)$ holonomy (which exists by **Theorem 11.9**).

Theorem 11.7 (Kodaira embedding theorem [28]) A compact complex manifold M is a **projective variety** (i.e. can be written as the zero-locus of finitely-many homogeneous polynomials in some $\mathbb{P}^N$) if and only if it admits a Kähler-Hodge metric, i.e. a Kähler metric whose Kähler form ω represents the Chern class of a holomorphic line bundle $\mathcal{L} \rightarrow M$, that is,

$$\omega \in H^2(M, \mathbb{Z}) \cap H^{1,1}(M) \quad (\text{a Hodge class}). \quad (11.47)$$

Corollary 11.8 All compact strict m -CYs with $m \geq 3$ are projective varieties.

Proof One has $h^{2,0} = 0$, so $H^2(X, \mathbb{C}) = H^{1,1}(X)$ and all harmonic forms have type (1, 1). X has a Kähler form ω which represents a non-zero class in $H^2(X, \mathbb{R})$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, for all $\epsilon > 0$ we may find a harmonic (1, 1) form ω' which represents a class in $H^2(X, \mathbb{Q})$ and $\|\omega' - \omega\|^2 < \epsilon$. For ϵ small enough, the closed (1, 1)-form ω' corresponds to a positive Kähler metric, since positivity is an open condition for metrics on X compact. A multiple of $[\omega']$ is an integral class. Now apply Kodaira embedding theorem. $\square$

Proposition 11.4 (Beauville [35]) X a CY in the sense (*). The fundamental group $\pi_1(X)$ is either finite or infinite. In the last case, there is a finite cover $X' = T^2 \times Y$ with T^2 a flat 2-torus and Y a Calabi–Yau manifold of dimension $\dim_{\mathbb{C}} X - 1$.

Proof A special case of **Corollary 11.3**. $\square$

Corollary 11.9 X a 3-CY. X strict $\Leftrightarrow \pi_1(X) < \infty$.

There are 14 smooth types of 3-CY's (in the sense $(*)$) which have infinite π_1 [36–38]. 6 of them are quotients of an Abelian variety by a finite group acting freely, and 8 are finite quotients of $E \times K$ with E an elliptic curve and K a K3 surface.

Calabi–Yau Manifolds of Complex Dimension 2

Since

$$SU(2) \simeq Sp(1), \quad (11.48)$$

in two complex dimensions being Calabi–Yau is equivalent to being hyperKähler, which in this case is equivalent to the Riemann tensor being anti-self-dual

$$R^i{}_j = - * R^i{}_j, \quad (11.49)$$

that is, a 2-CY is nothing else than a 4d gravitational instanton at zero cosmological constant. Non-compact examples of 2-CY are the Taub-NUT metrics to be discussed in Sect. 14.1. For a compact real 4-manifold M the Chern–Gauss–Bonnet and Hirzebruch signature theorems read [39]

$$\sum_i (-1)^i b_i = \chi = \frac{1}{32\pi^2} \int_M \epsilon_{abcd} R^{ab} \wedge R^{cd}, \quad (11.50)$$

$$b_2^+ - b_2^- = \tau = \frac{1}{48\pi^2} \int_M \epsilon_{abcd} R^{ab} \wedge * R^{cd}, \quad (11.51)$$

where b_2^+, b_2^- are (respectively) the number of self-dual and anti-self-dual harmonic 2-forms on M . From Eqs. (11.50), (11.51) we see that, whenever the Riemann curvature is anti-self-dual, we have

$$\chi = -3\tau/2. \quad (11.52)$$

The signature theorem (or the Lefschetz $sl(2)$ decomposition [20, 28]) yields for a compact Kähler 2-fold

$$\tau = 2h^{2,0} - h^{1,1} + 2. \quad (11.53)$$

For a strict 2-CY the only non-zero Hodge numbers are

$$h^{0,0} = h^{2,0} = h^{0,2} = h^{2,2} = 1, \text{ and } h^{1,1}, \quad (11.54)$$

and then

$$\chi = 4 + h^{1,1}, \quad (11.55)$$

so Eq. (11.52) becomes

$$4 + h^{1,1} = -\frac{3}{2}(4 - h^{1,1}) \Rightarrow h^{1,1} = 20, \quad \chi = 24, \quad \tau = -16. \quad (11.56)$$

Corollary 11.10 *A compact strict (i.e. non flat) Calabi–Yau 2-fold X_2 is simply-connected, $\pi_1(X_2) = 1$.*

Proof By **Proposition 11.4** $\pi_1(X_2)$ is finite. The universal cover $\tilde{X}_2$ is a finite cover, hence still a compact strict 2-CY. Now $\chi(\tilde{X}_2) = n \chi(X_2)$ where n is the degree of the universal cover. By Eq. (11.56) $n = 1$, i.e. $\tilde{X}_2 \equiv X_2$. $\square$

Since $SU(2)$ is simply-connected, X is a spin-manifold, so the second Stiefel–Whitney class $w_2(X) = 0$. Then, by Wu’s formula [40], the intersection form in $H^2(X, \mathbb{Z})$ is *even*, while it is self-dual by Poincaré duality. Thus $H^2(X, \mathbb{Z})$ is an *even, self-dual lattice of signature*

$$(2h^{2,0} + 1, h^{1,1} - 1) \equiv (3, 19). \quad (11.57)$$

By the classification of lattices (cf. Sect. 7.5) there is precisely *one* such lattice modulo equivalence.

Theorem 11.8 (Freedman [41, 42]) *For each even, self-dual lattice there is a **unique** compact simply-connected topological 4-manifold M with that lattice as the intersection form $H_2(M, \mathbb{Z}) \times H_2(M, \mathbb{Z}) \rightarrow \mathbb{Z}$.*

Corollary 11.11 *All compact strict 2-CY’s are homeomorphic.*

In fact one shows (Kodaira [43]) the stronger result that the compact strict 2-CY are all *diffeomorphic*, so we may view them as a unique smooth 4-manifold S_{smooth} which admits a family of inequivalent complex structures with trivial canonical bundle. The complex manifolds in this family are called *K3 surfaces*. An example of K3 surface is a smooth quartic hypersurface in $\mathbb{P}^3$. For an exhaustive treatment of the beautiful geometry of K3 surfaces see [44] and the literature cited therein. There are projective K3 as well as non-algebraic K3 (even K3 with no non-constant meromorphic functions). The moduli space of complex structures on K3 has complex dimension 20, while an algebraic family of K3 surfaces has just dimension 19.

Calabi–Yau Manifolds of Complex Dimension 3

Strict 3-CY have two independent Hodge numbers: $h^{1,1} = h^{2,2}$ and $h^{2,1} = h^{1,2}$; the other non-zero being $h^{0,0} = h^{3,0} = h^{0,3} = h^{3,3} = 1$. The Euler characteristic is

$$\chi \equiv \sum_{p,q} (-1)^{p+q} h^{p,q} = 2(h^{1,1} - h^{2,1}). \quad (11.58)$$

Several thousand topologically distinct 3-CY are known, see e.g. [45–47]. Various mathematical and physical considerations, including the swampland conjectures on quantum consistent theories of gravity [48], suggest the following:

Conjecture 11.1 *The number of topological types of 3-CY is finite.*

A bolder version of the conjecture states $h^{1,1} + h^{2,1} \leq 499$. A (simply-connected) compact smooth manifold which admits one complex structure with trivial canonical

bundle, typically admits several inequivalent such structures. There are, however, special smooth 6-manifolds, called *rigid* 3-CY, which admits a *unique* such complex structure. They are relatively “rare”: ≈ 50 examples of rigid 3-CY are known.

Families of m -CY Manifolds

Non-rigid CYs (in the sense $(*)$) form families with the same underlying real manifold but inequivalent complex structures. To study the complex deformations of the m -CY X_0 we consider any holomorphic fibration of complex manifolds

$$\varpi : X \rightarrow S, \quad S \text{ connected}, \quad (11.59)$$

whose fiber over the reference point $0 \in S$ is our m -CY X_0 . By general theory [2, 49] all fibers are diffeomorphic, see **BOX 11.1**

$$X_s \simeq_{\text{smooth}} X_0, \quad \forall s \in S \Rightarrow H^1(X_s, \mathbb{C}) = 0 \quad \forall s \in S. \quad (11.60)$$

BOX 11.1 — Families of Complex Manifolds and Monodromy

Consider the holomorphic family of complex m -folds $\{X_s\}_{s \in S}$, seen as a holomorphic fibration $\varpi : X \rightarrow S$ with X_s the fiber over s . Let s^i be local coordinates for S . The infinitesimal deformation $X_{s^i} \rightarrow X_{s^i + \delta s^i}$ is a vector field on X_{s^i} . We pack such fields into a 1-form $\mathcal{A}$ on S valued in the Lie algebra $C^\infty(X_s, T_{\mathbb{R}} X_s)$ of vector fields. Consider a path $\gamma \subset S$ with $\gamma(0) = s_0$, $\gamma(1) = s_1$. Integrating $\mathcal{A}$ along γ produces a diffeomorphism (cf. pp. 37–39 of [50])

$$F_\gamma : X_{s_0} \xrightarrow{\sim} X_{s_1},$$

which depends on γ . *This proves that all fibers are diffeomorphic.*

Two homotopic paths γ, γ' define homotopic maps $F_\gamma \sim F_{\gamma'}$, hence we have well-defined maps

$$\pi_1(S) \longrightarrow \text{MCG}(X_{s_0}) \longrightarrow \text{Aut}(H^m(X_{s_0}, \mathbb{Z})) \xlongequal{m \text{ odd}} Sp(b_m, \mathbb{Z})$$

here $\text{MCG}(X_{s_0})$ is the mapping class group and $\text{Aut}(H^\bullet(X_{s_0}, \mathbb{Z}))$ the automorphisms group of the cohomology which preserves the intersection form. The image Γ of $\pi_1(S)$ in $\text{Aut}(H^m(X_{s_0}, \mathbb{Z}))$ is the *monodromy group* of the family $X \rightarrow S$, while the induced map

$$\varrho : \pi_1(S) \rightarrow \text{Aut}(H^m(X_{s_0}, \mathbb{Z}))$$

is called the *monodromy representation*. It is independent of the base point s_0 modulo conjugacy.

Then, by the exponential exact sequence [29], we have the isomorphism¹²

¹² Recall from **BOX 1.6** that the Picard group $\text{Pic}(X)$ is the group of isomorphism classes of line bundles $L \rightarrow X$. For X smooth, it is also the group of divisors on X modulo linear equivalence.

$$\mathrm{Pic}(X_s) \simeq H^2(X_s, \mathbb{Z}) \cap H^{1,1}(X_s), \quad [L] \mapsto c_1(L), \quad (11.61)$$

so that the fiber's canonical bundle $\mathcal{K}_s$ is trivial iff $c_1(X_s) \equiv c_1(TX_s) = 0$. Since $c_1(X_0) = 0$ while $\mathcal{S}$ is connected, by continuity $c_1(X_s) = 0$ for all $s \in \mathcal{S}$ and all fibers X_s are strict Calabi–Yau. The Hodge numbers $h^{p,q}$ are constant in the family, indeed they can be written in terms of topological invariants as follows

$$h^{1,1} = \dim H^2(X_s), \quad h^{2,1} = \frac{1}{2} \dim H^3(X_s) - 1. \quad (11.62)$$

We say that the CY X_s is a *deformation* of the CY X_0 . From the smooth viewpoint X_s is the same manifold as X_0 , but it is endowed with an *inequivalent* complex structure. We say that the family $\mathcal{X} \rightarrow \mathcal{S}$ is *universal* if it contains all deformations of X_0 . The base $\mathcal{S}$ is then a cover of the actual complex moduli space $\mathcal{M}$. The basic result is

Theorem 11.9 (Yau [51]) *X_0 a compact m -CY. The underlying smooth manifold X admits a **unique** Riemannian metric of holonomy $SU(m)$ (a CY metric) for each complex structure $s \in \mathcal{M}$ and Kähler cohomology class $[\omega] \in H^2(X_0, \mathbb{R})$.*

Corollary 11.12 *The **classical** moduli space of Ricci-flat Kähler metrics on X is*

$$\mathcal{M} \times \mathcal{K}, \quad \mathcal{M} = \text{universal moduli space}, \quad \mathcal{K} = \text{Kähler cone} \quad (11.63)$$

The Kähler cone $\mathcal{K} \subset H^{1,1}(X, \mathbb{R})$ is the strict, open, convex cone¹³ which contains the classes $[\omega]$ of all positive-definite Kähler metrics on X .

The moduli space $\mathcal{M}$ has a rich geometric structure both locally and globally. Below we give a quick review of Kodaira–Spencer theory of deformations of complex structures. We already outlined it in **BOX 1.8** for the case of complex dimension 1.

Note 11.1 By general theory [2, 52] the moduli space $\mathcal{M}$ of (projective) Calabi–Yau manifolds is a quasi-projective algebraic variety. By Hironaka theorem [53] it can be written (non uniquely) as $\mathcal{M} = \overline{\mathcal{M}} \setminus D$, where $\overline{\mathcal{M}}$ is a (compact) smooth projective complex variety and D is an (effective) simple normal crossing divisor. All bounded pluri-subharmonic function¹⁴ on $\mathcal{M}$ is constant (Liouville property).

¹³ By definition $\mathcal{K}$ is a *convex cone* iff $\omega_1, \omega_2 \in \mathcal{K}$ implies $\lambda_1 \omega_1 + \lambda_2 \omega_2 \in \mathcal{K}$ for all $\lambda_1, \lambda_2 \in \mathbb{R}_{>0}$. It is *strict* if, in addition, $0 \neq \omega \in \mathcal{K}$ implies $-\omega \notin \mathcal{K}$.

¹⁴ A smooth function Φ on a complex manifold is *pluri-subharmonic* iff the matrix $\partial_j \partial_{\bar{k}} \Phi$ is non-negative [54]. In particular, a global Kähler potential (when it exists) is pluri-subharmonic.

11.1.4 Ultra-short Review of Kodaira–Spencer (KS) Theory

A general reference for KS theory is [49]; for a review in physicist’s style, see [7]. In this subsection M is a compact complex manifold, not necessarily Calabi–Yau.

Specifying a complex structure on a (even dimensional) smooth manifold X^{smooth} is equivalent to specifying which *local* smooth functions are holomorphic or, technically, to describe the sheaf $\mathcal{O}_X$ of germs of holomorphic functions as a subsheaf of the sheaf $\mathcal{A}$ of germs of smooth functions.¹⁵ The local holomorphic functions h satisfy the Cauchy–Riemann equation $\bar{\partial}h = 0$, i.e. are the smooth functions in the kernel of the Dolbeault operator $\bar{\partial}$. A family $\{X_s\}_{s \in S}$ of complex structures on a smooth manifold X^{smooth} is a family of Cauchy–Riemann equations of the form

$$\bar{\partial}_s h = 0 \quad \Leftrightarrow \quad h \text{ is holomorphic in complex structure } X_s, \quad (11.64)$$

where $\bar{\partial}_s$ is a 1st-order differential operator from smooth functions to 1-forms whose kernel is locally isomorphic to the ring of local holomorphic functions in $\mathbb{C}^m$. $\bar{\partial}_s$ kills the constants and satisfies the integrability condition $\bar{\partial}_s^2 = 0$.

We fix a complex structure on the central fiber X_0 of the family, and use its holomorphic coordinates as *smooth* coordinates for all X_s (this makes sense since X_s is diffeomorphic to X_0). Then, with no loss, we may write

$$\bar{\partial}_s = d\bar{z}^i (\partial_{\bar{z}^i} + \phi(s)_{\bar{i}}^j \partial_{z^j}) \equiv \bar{\partial} + \phi(s) \quad (11.65)$$

where¹⁶

$$\phi \equiv \phi(s) \in C^\infty(X_0, \Lambda^{0,1}(T^{1,0})) \quad (11.66)$$

is the *Kodaira–Spencer (KS) vector*, which satisfies the integrability condition

$$\bar{\partial}_s^2 = 0 \quad \Rightarrow \quad \bar{\partial}\phi + \frac{1}{2}[\phi, \phi] = 0 \quad (11.67)$$

called the KS equation. A ϕ satisfying the KS equation describes a deformation of the complex structure of X_0 . However not all such deformations are non-trivial. The moduli space of the complex structures on X^{smooth} is then the space of solutions to the KS equation modulo the natural “gauge-equivalence” given by global diffeomorphisms. The ones homotopic to the identity may be written as

$$\bar{\partial}_s \sim e^{-\xi} \bar{\partial}_s e^\xi, \quad \xi \in C^\infty(X_0, T^{1,0}), \quad (11.68)$$

¹⁵ With the condition that $(X, \mathcal{O}_X)$ is locally isomorphic to $(\mathbb{C}^n, \mathcal{O}_{\mathbb{C}^n})$ as a locally ringed space [55, 56], i.e. a complex manifold is locally modeled by $\mathbb{C}^n$ with its local holomorphic functions.

¹⁶ We write $\Lambda^{p,q}(\mathcal{V})$ for the space of smooth forms of type (p, q) on X_0 with coefficients in the holomorphic vector bundle $\mathcal{V}$ [22] and simply $T^{1,0}$ for the holomorphic tangent bundle $T^{1,0}X_0$.

which describe “pure gauge” solutions to the flatness equation (11.67). The space of solutions to Eq. (11.67) modulo the ones of the form (11.68) yield the *deformation space* $\mathcal{S}$ of the complex structure of X_0 . To get the actual moduli space $\mathcal{M}$ we have to mod out the action of the “large” diffeomorphisms which act via the mapping class group $\text{Mod}(X)$. Thus the *coarse* moduli space is¹⁷

$$\mathcal{M} = \mathcal{S}/\text{Mod}(X). \quad (11.69)$$

Below we shall need the following easy

Lemma 11.1 *In complex structure $s \in \mathcal{M}$ the fibers of the sub-bundle of $(1, 0)$ -forms, $\Lambda_s^{1,0} \subset T^*X \otimes \mathbb{C}$, are spanned by the forms*

$$\omega(s)^j = dz^j - \phi(s)_{\bar{k}}^j d\bar{z}^k \quad (11.70)$$

$\omega(s)^j$ depends holomorphically on the coordinates s^a of $\mathcal{M}$.¹⁸ A (p, q) -form on X_s , seen as a $(p + q)$ -form on the smooth manifold X_0 , has then the form

$$\Phi(s, \bar{s})_{j_1 \dots j_p \bar{l}_1 \dots \bar{l}_q} \omega(s)^{j_1} \wedge \dots \wedge \omega(s)^{j_p} \wedge \overline{\omega(s)^{\bar{l}_1}} \wedge \dots \wedge \overline{\omega(s)^{\bar{l}_q}}. \quad (11.71)$$

Proof $\Lambda^{1,0}$ is spanned by the differentials df of the holomorphic functions f which satisfy $\bar{\partial}_s f = d\bar{z}^i (\partial_{\bar{z}^i} f + \phi(s)_{\bar{i}}^j \partial_{z^j} f) = 0$. Hence

$$df = dz^i \partial_{z^i} f + d\bar{z}^j \partial_{\bar{z}^j} f = dz^i (\partial_{z^i} - d\bar{z}^j \phi(s)_{\bar{j}}^i \partial_{z^j} f) = \partial_{z^i} f (dz^i - \phi(s)_{\bar{j}}^i d\bar{z}^j) = (\partial_{z^i} f) \omega(s)^i.$$

□

Infinitesimal Deformations We consider infinitesimal deformations of X_0 . For infinitesimal KS vectors we replace Eqs. (11.67), (11.68) by their linearizations

$$\bar{\partial}\phi = 0, \quad \phi \sim \phi + \bar{\partial}\xi, \quad (11.72)$$

which identify the *formal* infinitesimal deformations with the cohomology space

$$H^1(X_0, T^{1,0}) \equiv \frac{\left\{ \ker \bar{\partial}: \Lambda^{0,1}(T^{1,0}) \rightarrow \Lambda^{0,2}(T^{1,0}) \right\}}{\left\{ \text{im } \bar{\partial}: C^\infty(T^{1,0}) \rightarrow \Lambda^{0,1}(T^{1,0}) \right\}} \quad (11.73)$$

cf. Beltrami differentials in one complex dimension, i.e. the zero-modes of the b ghost. However in dimension > 1 the result (11.73) is purely formal: a would-be infinitesimal deformation $\phi_1 \in H^1(X_0, T^{1,0})$ may be *obstructed* in higher order. Consider the expansion of the KS vector in the small parameter ϵ

¹⁷ **Warning.** Here we are oversimplifying; for a precise treatment of m -CY, see [57–59].

¹⁸ For more details see e.g. [7].

$$\phi(\epsilon) = \epsilon\phi_1 + \epsilon^2\phi_2 + \cdots, \quad (11.74)$$

and solve the KS equation (11.67) order by order in ϵ

$$\bar{\partial}(\epsilon\phi_1 + \epsilon^2\phi_2 + \cdots) = \frac{1}{2}\epsilon^2[\phi_1, \phi_1] + \cdots. \quad (11.75)$$

To order ϵ^1 we get the condition $\phi_1 \in H^1(X_0, T^{1,0})$. To be able to go on, and solve the equation to the next order for ϕ_2 , we need the first term in the RHS, $[\phi_1, \phi_1]$, to be $\bar{\partial}$ -exact, while in general it is only $\bar{\partial}$ -closed. When $[\phi_1, \phi_1]$ is not zero in $H^2(X_0, T^{1,0})$, the formal infinitesimal deformation ϕ_1 cannot be completed to an actual deformation: we say that it is *obstructed* at higher order. Hence the tangent space to the universal deformation space at the central fiber, $T_0\mathcal{S}$ in general is *smaller* than $H^1(X_0, T^{1,0})$. The only thing that we may state in general is

$$\dim T_0\mathcal{S} \leq \dim H^1(X_0, T^{1,0}), \quad (11.76)$$

with the proviso that the deformation space *needs not* to be smooth at $s = 0$. This is the situation with the deformations of *general* complex manifolds. Luckily Calabi–Yau behave much better in this respect, as we are going to show.

Complex Deformations: the Calabi–Yau Case

We specialize to the case where X_0 is a m -CY. The canonical bundle, $\mathcal{K} \equiv \wedge^m \Lambda^{1,0}$, is trivial and hence

$$\begin{aligned} \wedge^k T^{1,0} &\simeq \wedge^{m-k} \Lambda^{1,0} \Rightarrow \\ &\Rightarrow H^1(X_0, T^{1,0}) \simeq H^1(X_0, \Lambda^{m-1,0}) \equiv H^{m-1,1}(X_0), \end{aligned} \quad (11.77)$$

and we may identify (formal) infinitesimal deformations of complex structure $\phi \in H^1(X, TX)$ with harmonic $(m-1, 1)$ -forms. Using this identification one shows the following.

Theorem 11.10 (Tian [60]) *All infinitesimal deformations of a compact m -CY are unobstructed and the universal deformation space $\mathcal{S}$ is a smooth complex manifold of dimension $h^{m-1,1}$.*¹⁹

By construction the KS vector

$$\phi(s) = s^a \phi_a + O(s^2) \quad \{\phi_a\} \text{ a basis of } H^1(X, T^{1,0}) \quad (11.78)$$

depends *holomorphically* on the deformation parameters s^a .

¹⁹ The moduli space $\mathcal{M}$, Eq. (11.69), has orbifold singularities. E.g. the 1-CY are elliptic curves, $\mathcal{S}$ is the upper half-plane $\mathcal{H}$, which is smooth, while $\mathcal{M}_1 = \mathcal{H}/SL(2, \mathbb{Z})$ has two orbifold points i , $e^{2\pi i/3}$ fixed by (conjugacy classes of) finite subgroups of the modular group.

Corollary 11.13 (Local Torelli) *The map which associates to a point $s \in \mathcal{M}$ the Hodge decomposition $H^m(X_0, \mathbb{C}) \equiv H^m(X_s, \mathbb{C}) = \oplus_{p,q} H^{p,q}(X_s)$ is **locally injective**.*

Indeed, an element $\phi_s \in T_s \mathcal{M}$ is a KS vector. Contraction gives a map $\phi_s : H^{m,0}(X_s) \rightarrow H^{m-1,1}(X_s)$, so $T_s \mathcal{M} \subseteq H^{m-1,1}(X_s) \otimes H^{m,0}(X_s)^\vee$, and since the two spaces have the same dimension

$$T_s \mathcal{M} \simeq H^{m-1,1}(X_s) \otimes H^{m,0}(X_s)^\vee. \quad (11.79)$$

Note 11.2 From **Theorem 11.10**, it is clear that the deformation space $\mathcal{S}$ is contractible. However the moduli space $\mathcal{M}$ is topologically very non-trivial as it is obvious from *Note 11.1*.

Natural Metric on the Complex Moduli

On $\mathcal{M}$ we have a natural Kähler metric: the *Weil–Petersson* (WP) one [60]

$$G_{a\bar{b}} = \int_X g_{i\bar{j}} g^{\bar{k}l} \partial_{s^a} \phi_k^i \partial_{\bar{s}^b} \bar{\phi}_l^{\bar{j}} \sqrt{g} d^{2m} z. \quad (11.80)$$

Let $\Omega(s)$ be a (local) holomorphic m -form in the total space $\mathcal{X}$ which restricts to a non-zero $(m, 0)$ holomorphic form on the CY fibers. $\Omega(s)$ is unique up to multiplication by a (local) holomorphic function $\Omega(s) \rightarrow f(s)\Omega(s)$. From **Lemma 11.1** $\Omega(s)$ has the general structure

$$\Omega(s) = F(s) \epsilon_{i_1 \dots i_m} \omega(s)^{i_1} \wedge \dots \wedge \omega(s)^{i_m}. \quad (11.81)$$

Then, for some functions f_a ,

$$\partial_{s^a} \Omega(s) = -(\partial_{s^a} \phi) \rfloor \Omega(s) + f_a(s) \Omega(s) \quad (11.82)$$

$$\Rightarrow G_{a\bar{b}} = -\partial_{s^a} \partial_{\bar{s}^b} \log \left(-i \int_X \Omega(s) \wedge \overline{\Omega(s)} \right). \quad (11.83)$$

Equation (11.83) shows that *the Weil–Petersson metric on $\mathcal{M}$ is Kähler*.

Note 11.3 One unpleasant aspect of the WP metric is that $\mathcal{M}$ is *non-complete* when endowed with this metric [61]. On $\mathcal{M}$ there is a mathematically nicer Kähler metric, the *Hodge* one [2, 57], which *is* complete and has other nice properties (e.g. its holomorphic bi-sectional curvatures are non-positive). For its physical meaning see [62] and references therein.

11.2 Superstrings on CY Manifolds: The World-Sheet Perspective

We recall that, given *any* Riemannian manifold (M, g) , we can define a (1,1)-supersymmetric 2d σ -model with Lagrangian [7]

$$\mathcal{L}_\sigma = -\frac{1}{2}g_{ij}\partial_\mu\phi^i\partial^\mu\phi^j + \frac{i}{2}g_{ij}\bar{\psi}^i\gamma^\mu\nabla_\mu\psi^j + \frac{1}{12}R_{ijkl}\bar{\psi}^i\psi^k\bar{\psi}^j\psi^\ell. \quad (11.84)$$

The connection ∇ is the pull-back to the world-sheet of the target Levi-Civita one

$$\nabla_\mu\psi^i = \partial_\mu\psi^i + \partial_\mu\phi^j\Gamma_{jk}^i\psi^k. \quad (11.85)$$

The Lagrangian (11.84) enhances from (1, 1) to (2, 2) SUSY iff the target manifold M is Kähler, and to (4, 4) SUSY iff M is hyperKähler [7]. A further SUSY enhancement requires M to be *flat* of suitable dimension: this is the rigid SUSY counterpart to the statement that in SUGRA with more than 8 SUSY the scalar target space must be a locally symmetric space,²⁰; cf. Sect. 8.1.

Claim 11.1 If the manifold M admits a CY metric (with Kähler form ω_{CY}), then it also admits a Kähler metric (with Kähler form ω) such that the 2d SUSY σ -model is *exactly* (i.e. non-perturbatively) (2, 2) superconformal.

Note 11.4 There is NO general claim that $\omega = \omega_{CY}$ although this equality is asymptotically valid as $\hbar \rightarrow 0$. There are some higher order corrections to the CY metric at four-loops [63]. Equality holds for the special case that the CY is actually HK.

Proof For a general SUSY σ -model, with g_{ij} an arbitrary Riemannian metric, a fixed point of the Renormalization Group (RG) requires that the exact (non-perturbative) β -function for the metric (seen as a 2d coupling) satisfies the condition

$$\mu \frac{\partial g_{ij}}{\partial \mu} \equiv \beta_{ij} = \mathcal{L}_v g_{ij} \equiv \nabla_i v_j + \nabla_j v_i \quad (11.86)$$

for some *complete* vector field v^i . If this equation holds, the RG flow acts on the Riemannian manifold by a diffeomorphisms, that is, by a mere field redefinition which does not affect the physics. In a given renormalization scheme, this implies the existence of suitable higher-loop counter-terms which we may use to modify order by order the target metric to get $\beta_{ij} = 0$.

β_{ij} is a symmetric tensor which has a universal (i.e. model independent) expression in terms of the local invariants of the Riemannian geometry of M , i.e. the Riemann tensor and its covariant derivatives of all orders [64]. To one-loop β_{ij} must be proportional to the Ricci tensor R_{ij} , since this is the only symmetric tensor with the appropriate scaling property under Weyl rescaling. If M admits a CY metric, it is in particular complex and Kähler, and these two conditions are preserved by the RG flow, since RG preserves the Poincaré (2,2) SUSY which requires M to be Kähler. Hence, for all (2,2) σ -model, the β -function may be seen as a real (1, 1)-form on M

$$\beta \equiv i\beta_{i\bar{j}}dz^i \wedge d\bar{z}^{\bar{j}} \equiv i\left(\mu \frac{\partial g_{i\bar{j}}}{\partial \mu}\right)dz^i \wedge d\bar{z}^{\bar{j}} \equiv \frac{\partial}{\partial \log \mu} \omega \quad (11.87)$$

²⁰ Indeed a locally symmetric space which is Ricci-flat is flat.

where ω is the Kähler form. Now the RG flow is an equation for $(1, 1)$ -forms on M , and since the RHS of (11.87) is d -closed so must be its LHS:

$$d\beta = \partial\beta = \bar{\partial}\beta = 0. \quad (11.88)$$

Written in terms of differential forms, the fixed point condition (11.86) then becomes

$$\beta = \mathbb{L}_v \omega = d(i_v \omega) \quad (11.89)$$

where we used the Cartan identity $\mathbb{L}_v = di_v + i_v d$ for the Lie derivative $\mathbb{L}_v$ acting on forms, and the fact that the Kähler form is closed $d\omega = 0$. Since the map $v \mapsto i_v \omega$ is an isomorphism between the spaces of vector fields and 1-forms, we get that a $(2, 2)$ fixed point of the RG flow (in the sense (11.86)) exists iff $\beta = 0$ in cohomology. By the $\partial\bar{\partial}$ -Poincaré lemma [28], the condition for a RG fixed point is also equivalent to $\beta = i \partial\bar{\partial}K$ for a global smooth real function K on M . The function K can be used to construct explicit 2d local counter-terms which kill the trace anomaly. To make the situation more transparent, we present a variant of the argument which may be more familiar to the reader.

Alternative proof: Relation with the Adler–Bardeen theorem In a $(2, 2)$ σ -model we have a classically conserved $U(1)$ axial current $j_\mu^A = g_{i\bar{j}} \bar{\psi}^{\bar{j}} \gamma_3 \gamma_\mu \psi^i$ because the model is *classically* $(2, 2)$ superconformal, and this $U(1)$ is the axial R-current of the $\mathcal{N} = 2$ superconformal algebra; cf. Sect. 2.10. This current belongs to the same superfield as $T_{\mu\nu}$, and its anomaly belongs to the same superfield as the trace anomaly²¹

$$d * j^A = \phi^* \beta. \quad (11.90)$$

The Adler–Bardeen theorem [65, 66] states that, if the axial $U(1)$ anomaly vanishes at one-loop, it vanishes to all loops in the sense that there exist counter-terms which kill it. For a CY σ -model the one-loop anomaly $\phi^* \beta|_{1\text{-loop}} = 0$, hence $\beta = 0$ to all loops in a suitable renormalization scheme.

Note 11.5 For yet another proof, see [67].

□

As already mentioned, the SCFT metric is slightly different from the CY one, due to higher quantum corrections [63]. This does not happen when the target metric satisfies the stronger condition of being hyperKähler: the σ -model has a $SU(2)_R$ global symmetry which rotates the 3-Kähler forms in the adjoint representation; since this global non-anomalous symmetry is preserved by the RG flow, the metric remains hyperKähler all along the flow, while neither the complex structure nor the Kähler class gets renormalized because a HK manifold is automatically Ricci-flat. Then, by Yau uniqueness, the HK metric does not flow. This is the 2d $(4, 4)$ SUSY non-renormalization theorem.

In conclusion: spacetime geometries of the form $\mathbb{R}^{d-1,1} \times M$ with M a CY manifold of complex dimension $5 - d/2$ are *exact* backgrounds for the Type IIA/IIB superstring invariant under d -dimensional Poincaré symmetry, that is, they describe d -dimensional vacuum states of superstring theory. We shall show momentarily that they preserve spacetime $\mathcal{N} = 2$ SUSY. Later we shall describe how they give compactifications of the $E_8 \times E_8$ heterotic string preserving 4d $\mathcal{N} = 1$ SUSY.

²¹ Through this chapter we see the bosonic configurations of the σ -model as maps $\phi: \Sigma \rightarrow M$.

11.2.1 Calabi–Yau 2d σ -Models as (2, 2) SCFTs

To avoid special cases, we assume M to be a *strict* CY with $\dim_{\mathbb{C}} M \geq 3$. Then $h^{2,0} = 0$, all harmonic 2-forms have type (1, 1), and M is a smooth projective variety.

Conformal Manifold

We study the conformal manifold $\mathcal{C}$ of the (2,2) SCFT at the fixed point of the SUSY CY σ -model. $\mathcal{C}$ is equipped with a natural Riemannian metric, the Zamolodchikov one [68]. We may add to the action a coupling to a target-space 2-form field B_{ij}

$$S = \int_{\Sigma} \mathcal{L}_{\sigma} + \int_{\Sigma} \phi^* B. \quad (11.91)$$

If B is *flat*, i.e. its field-strength vanishes, $H \equiv dB = 0$, this coupling is a generalized θ -angle, coupled to a topological invariant of the field configuration $\phi: \Sigma \rightarrow M$, which does not spoil conformal invariance. On the other hand, if the 2-form B is *d*-exact, $B = d\xi$, the topological coupling is trivial. Thus the number of actual marginal deformations of the SCFT which arise from the B -term is $\dim H^2(M) \equiv h^{1,1}$. From Yau theorem we see that, in the limit of *weak coupling* $\hbar \rightarrow 0$, i.e. in the regime where classical geometry is a reliable description of the dynamics, the (2, 2) conformal manifold is asymptotic to the product space

$$\mathcal{M} \times \mathcal{K}_{\text{class}} \quad \text{as } \hbar \rightarrow 0, \quad (11.92)$$

of the complex moduli $\mathcal{M}$ and the *classical complexified* Kähler moduli $\mathcal{K}_{\text{class}}$

$$\mathcal{K}_{\text{class}} \equiv \{B + i\omega \in H^{1,1}(X) \mid \omega \in (\text{Kähler cone})\} \quad (11.93)$$

of complex dimension $h^{2,1}$ and $h^{1,1}$, respectively.²² At the quantum level, the conformal manifold gets corrected with respect to the classical geometry (11.92). Luckily the conformal manifold enjoys remarkable non-renormalization properties:

Claim 11.2 The (2, 2) conformal manifold $\mathcal{C}$ has the properties:

- (1) the *exact* (2,2) quantum conformal manifold is still a Riemannian product²³ of the complex moduli and (quantum) Kähler moduli spaces, $\mathcal{C} = \mathcal{M} \times \mathcal{K}$;
- (2) the complex moduli space $\mathcal{M}$ and its metric get **no** quantum correction: the classical geometric expressions are quantum exact. The *exact* Zamolodchikov metric of the factor $\mathcal{M}$ is the WP metric (11.80) on the complex moduli of M .

²² Complex manifolds of the form $\{Z \in \mathbb{C}^n : \text{Im } Z \in V\}$ with $V \subset \mathbb{R}^n$ a strict, convex cone are called *tube domains* or *Siegel domains of the first kind* [69].

²³ By a *Riemannian product* we mean a product $M_1 \times M_2$ of manifolds equipped with the metric $ds^2 = ds_1^2 + ds_2^2$ where ds_a^2 is a Riemannian metric on M_a ($a = 1, 2$).

Proof We will be sketchy; for more details see [70]. In the $(2, 2)$ superspace we have 4 odd coordinates, 2 on the left $\theta_+, \bar{\theta}_+$ and 2 on the right $\theta_-, \bar{\theta}_-$, and hence 4 odd super-derivatives $D_+, \bar{D}_+, D_-$ and $\bar{D}_-$. The chiral (anti-chiral) superfields X^i ($\bar{X}^i$) satisfy

$$\bar{D}_+ X^i = \bar{D}_- X^i = 0, \quad D_+ \bar{X}^i = D_- \bar{X}^i = 0. \quad (11.94)$$

The usual F-term couplings are

$$\int d^2 z d\theta_+ d\theta_- W(X^i), \quad \int d^2 z d\bar{\theta}_+ d\bar{\theta}_- \bar{W}(\bar{X}^i) \quad (11.95)$$

with $W(X^i)$ a holomorphic function, the *superpotential*. In 2d, we can also define *twisted chiral* (anti-chiral) superfields Σ^a (resp. $\bar{\Sigma}^a$)

$$\bar{D}_+ \Sigma^a = D_- \Sigma^a = 0, \quad D_+ \bar{\Sigma}^a = \bar{D}_- \bar{\Sigma}^a = 0, \quad (11.96)$$

and correspondingly twisted F-terms

$$\int d^2 z d\theta_+ d\bar{\theta}_- \tilde{W}(\Sigma^a), \quad \int d^2 z d\bar{\theta}_+ d\theta_- \tilde{\bar{W}}(\bar{\Sigma}^a) \quad (11.97)$$

with $\tilde{W}(\Sigma^a)$ a second holomorphic function, the “*twisted superpotential*”. Chiral and twisted-chiral superfields can be coupled in a SUSY-invariant way only in the D-terms

$$\int d^2 z d^4 \theta K(X^i, \bar{X}^j, \Sigma^a, \bar{\Sigma}^b) \quad (11.98)$$

which are both *irrelevant* in the IR (on dimensional grounds) and cohomologically trivial in the SUSY sense [4]. Marginal deformations of the action are either F-terms or twisted F-terms. The tangent space to the conformal manifold $\mathcal{C}$ then splits into chiral and twisted-chiral directions

$$TC = (\text{chiral marginal deformations}) \oplus (\text{twisted-chiral marginal deformations}). \quad (11.99)$$

SUSY implies that this splitting of TC is preserved by parallel transport with the Zamolodchikov Levi-Civita connection of $\mathcal{C}$. We go to the universal cover $\tilde{\mathcal{C}}$ of $\mathcal{C}$ and use **Theorem 11.4**, to conclude that $\tilde{\mathcal{C}}$ is globally a Riemannian product of two manifolds²⁴

$$\tilde{\mathcal{C}} = \mathcal{M} \times \mathcal{K} \quad (11.100)$$

spanned, respectively, by the chiral and the twisted-chiral marginal deformations

$$T\mathcal{M} = (\text{chiral marginal}) \quad T\mathcal{K} = (\text{twisted-chiral marginal}). \quad (11.101)$$

We claim that complex moduli are chiral deformations while Kähler moduli are twisted-chiral.²⁵ Since the splitting (11.100) is exact for all couplings, it suffices to check the claim at weak coupling, where geometry is reliable. Then a harmonic KS vector ϕ defines a superconformal chiral operator while a harmonic $(1, 1)$ -form ω a twisted-chiral operator, both with $(h, \bar{h}) = (1, 1)$, i.e. marginal

$$\bar{\phi}_i^{\bar{k}} g_{\bar{k}j} \psi_+^i \psi_-^j \quad \text{and, respectively,} \quad \omega_{i\bar{j}} \bar{\psi}_-^{\bar{j}} \psi_+^i. \quad (11.102)$$

²⁴ We are cheating a little bit: $\tilde{\mathcal{C}}$ is non-complete in the Zamolodchikov metric G_{ij} . Perturbing G_{ij} by its Ricci tensor, $G_{ij} \rightarrow G_{ij} + \epsilon R_{ij}$, we get a metric satisfying the conditions of the **Theorem**.

²⁵ We stress that the distinction *chiral* vs. *twisted-chiral* is a matter of conventions.

This proves part (1) of **Claim 11.2**. To show part (2) we reintroduce $\hbar$. The action, including the kinetic terms, is multiplied by $\hbar^{-1}$, so the effective σ -model Kähler class is $[\omega/\hbar]$. Thus $1/\hbar$ is the overall Kähler modulus, controlling the size of the CY manifold M . Then the deformation $\hbar \rightarrow \hbar + \delta\hbar$ is a twisted-chiral perturbation, i.e. a vector field $\partial_{\hbar} \in C^\infty(\mathcal{K}, T\mathcal{K})$ in the factor $\mathcal{K}$ of the quantum conformal manifold C . Taking the classical limit $\hbar \rightarrow 0$ we go to infinity along an integral curve of the vector $\partial_{\hbar}$ in the factor $\mathcal{K}$ of the conformal manifold $\mathcal{M} \times \mathcal{K}$; this limit does not affect the geometry of the first factor $\mathcal{M}$, which is then totally independent of $\hbar$. This entails that the classical geometry of $\mathcal{M}$ is *exact* at the quantum level. $\square$

Chiral Rings, Spectral Flow, and Spin-Fields

In this paragraph we study some important property of *abstract* 2d (2,2) SCFTs which may or may not have a Lagrangian realization as a σ -model. The conformal manifold is still a product of two spaces parametrized respectively by chiral and twisted-chiral marginal deformations: $C = \mathcal{M} \times \mathcal{K}$. Recall from Sect. 2.10 that the left-moving chiral currents of the 2d (2, 2) superconformal algebra are the energy–momentum $T_B(z)$, two supercurrents $G^\pm(z)$, and a $U(1)$ current $J(z)$ with OPEs (we set $\hat{c} \equiv c/3$)

$$J(z) G^\pm(w) \sim \pm \frac{1}{z-w} G^\pm(w), \quad J(z) J(w) \sim \frac{\hat{c}}{(z-w)^2}. \quad (11.103)$$

In addition we have right-moving chiral currents $\tilde{T}(\bar{z})$, $\tilde{G}^\pm(\bar{z})$, and $\tilde{J}(\bar{z})$. We focus on the left-moving side. The anti-commutators of the supercurrent modes $G_r^\pm$ are

$$\{G_r^+, G_s^-\} = 2L_{r+s} + (r-s)J_{r+s} + \hat{c} \left(r^2 - \frac{1}{4} \right) \delta_{r,-s}, \quad (11.104)$$

$$\{G_r^+, G_s^+\} = \{G_r^-, G_s^-\} = 0, \quad (11.105)$$

where $r, s \in \mathbb{Z}$ in the R-sector and $r, s \in \mathbb{Z} + \frac{1}{2}$ in the NS sector. In particular:

$$\{G_{-1/2}^+, G_{1/2}^-\} = 2L_0 - J_0, \quad \{G_{1/2}^+, G_{-1/2}^-\} = 2L_0 + J_0, \quad (11.106)$$

$$\{G_0^+, G_0^-\} = 2L_0 - \hat{c}/4 \quad \text{where } \hat{c} \equiv c/3. \quad (11.107)$$

In an unitary theory the LHS are non-negative operators: we get the following unitarity bounds on the Virasoro weight h and $U(1)$ charge q

$$\text{NS sector:} \quad h \geq \frac{1}{2}|q|, \quad (11.108)$$

$$\text{R sector:} \quad h \geq \frac{1}{8}\hat{c}. \quad (11.109)$$

States/operators which saturate these bound have special names [27, 70]:

- NS operators with $h = q/2 \geq 0$ are called *chiral primaries* c ;
- NS operators with $h = -q/2 \geq 0$ are called *anti-chiral primaries* a . The Hermitian conjugate of a chiral primary is an anti-chiral primary;

- the R sector states with $h = \hat{c}/8$ are called *Ramond ground states*: they are the SUSY vacua of the SCFT quantized on a circle S^1 with periodic fermions.

States/operators which saturate the NS/R unitary bounds (11.108), (11.109) belong to *short* superconformal multiplets, and hence are SUSY protected against quantum corrections by the usual arguments; cf. Chap. 8.

The 4 Chiral Rings Let $\phi_1(z), \phi_2(w)$ be two chiral primaries of $U(1)$ charges q_1 and q_2 , respectively. Their OPE has the form

$$\phi_1(z) \phi_2(w) = \sum_n (z-w)^{h_n-h_1-h_2} \Phi_n(w), \quad (11.110)$$

while the dimension h_n of the operator Φ_n satisfies (cf. (11.108))

$$h_n \geq \frac{1}{2} q_n \equiv \frac{1}{2} (q_1 + q_2) \equiv h_1 + h_2. \quad (11.111)$$

Thus the product (11.110) is *regular* as $z \rightarrow w$ and

$$\phi_1 \phi_2(w) \stackrel{\text{def}}{=} \lim_{z \rightarrow w} \phi_1(z) \phi_2(w), \quad (11.112)$$

if *non-zero*, is also a chiral primary saturating the equality $h = q/2$. We conclude that chiral primaries form a *ring* under the natural product (11.112). The same result holds for the anti-chiral primary operators. Taking into account the right-movers, in the NS-NS sector we have four rings of SUSY protected operators

$$(c, \tilde{c}), \quad (a, \tilde{a}), \quad (c, \tilde{a}), \quad (a, \tilde{c}) \quad (11.113)$$

called, respectively, the *chiral*, *anti-chiral*, *twisted-chiral* and *anti-twisted-chiral* ring. We stress that the difference between these four rings is purely conventional: flipping the (conventional) overall signs of the two currents $J(z) \leftrightarrow -J(z)$, $\tilde{J}(\bar{z}) \leftrightarrow -\tilde{J}(\bar{z})$ we transform one into the other. The chiral (resp. twisted-chiral) primaries are the first components of chiral (resp. twisted-chiral) superfields in the sense of Eqs. (11.94), (11.96). On the contrary, chiral/twisted-chiral superfields need not to be primary.²⁶ The operator 1 is the identity in all four protected rings (11.113).

From the anti-commutator

$$\{G_{3/2}^{\pm}, G_{-3/2}^{\mp}\} = 2L_0 \pm 3J_0 + 2\hat{c} \quad (11.114)$$

²⁶ Recall from Chap. 2 the notion of superconformal superfields: not all superfields are superconformal superfields. The chiral primaries are first components of superconformal superfields. More specifically, a general NS chiral state $|\psi\rangle$ satisfies $G_{-1/2}^+ |\psi\rangle = 0$, while a *primary* chiral state satisfies $G_{+1/2}^- |\psi\rangle = G_{-1/2}^+ |\psi\rangle = 0$. A non-zero state of the form $G_{-1/2}^+ |\eta\rangle$ is a chiral *descendent*.

we get a second unitarity bound in the NS sector

$$h - \frac{3}{2}|q| + \hat{c} \geq 0 \quad (11.115)$$

BOX 11.2 —Frobenius Algebras

The *Frobenius algebras* are a special class of—not necessarily (graded)commutative—finite-dimensional, associative, unital $\mathbb{C}$ -algebras with remarkable properties and a deep theory (see e.g. [71]). An algebra A is Frobenius iff the sub-category $\text{proj } A \subset \text{mod } A$ of its projective (right) modules coincides with the sub-category $\text{inj } A \subset \text{mod } A$ of its injective modules. Alternatively, it can be defined as an algebra that satisfies any one of the conditions in the following^a

Theorem (Brauer, Nesbitt, Nakayama). Let A be a $\mathbb{C}$ -algebra. The following are equivalent:

- (1) There exists a non-degenerate $\mathbb{C}$ -bilinear form $(-, -): A \times A \rightarrow \mathbb{C}$ such that $(a, bc) = (ab, c)$
- (2) There exists a linear form $\Phi: A \rightarrow \mathbb{C}$ such that $\ker \Phi$ does not contain a non-zero right ideal
- (3) There exists an isomorphism $\theta: A_A \rightarrow D(A)_A$ of right A -modules
- (4) There exists a linear form $\Phi' \rightarrow \mathbb{C}$ such that $\ker \Phi'$ does not contain a non-zero left ideal
- (5) There exists an isomorphism $\theta: {}_A A \rightarrow {}_A D(A)$ of left A -modules.

The Frobenius algebra is *symmetric* if the bilinear form is symmetric, i.e. if the form Φ is a “trace” $\Phi(ab) = \Phi(ba)$ for all $a, b \in A$. All commutative Frobenius algebras are symmetric.

Remark 11.1 There is a far-reaching generalization: the *Frobenius categories*; cf. Sect. 8.6 of [72].

^a **Notation.** A_A (resp. ${}_A A$) denotes the algebra A seen as a right (resp. left) module over itself. $D(-) = \text{Hom}_{\mathbb{C}}(-, \mathbb{C})$ is the “classical” duality in the category of vector spaces over $\mathbb{C}$. D maps right modules into left modules and left ones and viceversa. One has $D^2 = \text{Id}$.

which when applied to (anti-)chiral primaries, $h = |q|/2$, gives the bound

$$|q| \leq \hat{c}. \quad (11.116)$$

We conclude that in a *non-degenerate* SCFT, i.e. a model with a unique normalizable SL_2 -invariant vacuum and a *discrete* spectrum of CFT weights h , the following holds

Fact 11.3 *The 4 SUSY protected rings (11.113) are finite-dimensional, associative, super-commutative,²⁷ unital $\mathbb{C}$ -algebras, in facts Frobenius algebras (Box 11.2).*

²⁷ *Super-commutative* means that the elements of the ring are distinguished in bosonic and fermionic; the bosonic ones commute and the fermionic ones anti-commute. *Unital* means that the algebra contains an identity element 1.

We stress that **Fact 11.3** holds for *all* regular (2,2) supersymmetric QFT whether conformal or not [4, 70]. Indeed, the 4 rings were defined in terms of representations of Poincaré (2, 2) SUSY. When the (2, 2) QFT is actually a SCFT, the 4 protected rings have additional properties. We focus on the chiral ring $\mathcal{R} \equiv (\mathbf{c}, \tilde{\mathbf{c}})$, the story for the other rings being identical.

The ring $\mathcal{R}$ of a SCFT is graded by the $U(1) \times \tilde{U}(1)$ charges $(q, \tilde{q})$

$$\mathcal{R} = \bigoplus_{(q, \tilde{q}) \geq (0, 0)} \mathcal{R}_{(q, \tilde{q})}, \quad \mathcal{R}_{(q, \tilde{q})} \cdot \mathcal{R}_{(q', \tilde{q}')} \subseteq \mathcal{R}_{(q+q', \tilde{q}+\tilde{q}')}. \quad (11.117)$$

The only chiral primary with $(q, \tilde{q}) = (0, 0)$ is the identity, i.e. $\mathcal{R}_{(0, 0)} = \mathbb{C} \cdot \mathbf{1}$. Hence in the SCFT case the chiral ring $\mathcal{R}$ is a *graded local* $\mathbb{C}$ -algebra, that is, it has a unique maximal ideal $\mathfrak{m} = \bigoplus_{(q, \tilde{q}) \neq (0, 0)} \mathcal{R}_{(q, \tilde{q})}$ with residue field $\mathcal{R}/\mathfrak{m} = \mathbb{C}$.

Remark 11.2 By sake of comparison, we mention that in the opposite extremum, that is, in a *gapped* 2d (2,2) QFT, the chiral ring $\mathcal{R}$ is a *semi-simple* $\mathbb{C}$ -algebra [5], hence (if commutative) a product of copies of $\mathbb{C}$ [73].

Let $\{\phi_j\}_{j=0}^k$ be a basis of $\mathcal{R}$ with $\phi_0 = 1$. The product in $\mathcal{R}$ is

$$\phi_i \phi_j = C_{ij}^k \phi_k \quad (11.118)$$

for some coefficients C_{ij}^k . The matrices $(C_i)_j^k \equiv C_{ij}^k$ give a representation of the associative, graded-commutative ring $\mathcal{R}$. They depend on the couplings, i.e. on the point on the manifold $\mathcal{C}$. More precisely, C_{ij}^k is a *holomorphic* section of the bundle

$$(\odot^2 \mathcal{R}^\vee) \otimes \mathcal{R} \rightarrow \mathcal{M}. \quad (11.119)$$

Indeed by **Claim 11.2(1)** the coefficients C_{ij}^k are independent of the Kähler moduli. The proof that C_{ij}^k depends holomorphically on $\mathcal{M}$ is given in *Note 11.7* below.

The protected rings have a simple relation with marginal deformations

$$T_{\mathbb{R}} \mathcal{M} \otimes \mathbb{C} \simeq (\mathbf{c}, \tilde{\mathbf{c}})_{(1, 1)} \oplus (\mathbf{a}, \tilde{\mathbf{a}})_{(-1, -1)} \quad (11.120)$$

$$T_{\mathbb{R}} \mathcal{K} \otimes \mathbb{C} \simeq (\mathbf{c}, \tilde{\mathbf{a}})_{(1, -1)} \oplus (\mathbf{a}, \tilde{\mathbf{c}})_{(-1, 1)} \quad (11.121)$$

where $(\cdots)_{(q, \tilde{q})}$ stands for the subspace of charges $(q, \tilde{q})$ of the $U(1) \times \tilde{U}(1)$ -graded ring (cf. (11.117)). The first isomorphism is given by the map

$$\phi \in (\mathbf{c}, \tilde{\mathbf{c}})_{(1, 1)} \longmapsto \epsilon \int_{\Sigma} d^2 z \, d^2 \theta \, \phi, \quad (11.122)$$

together with its conjugate. To see that the RHS is exactly marginal, compare with the BRST-invariant vertices on the superstring (which describe marginal deformations). The second isomorphism is the twisted version of the first one.

Corollary 11.14 (1) *The two factors of the conformal manifold $\mathcal{C}$ carry a natural complex structure such that the holomorphic tangent bundle of $(1, 0)$ vectors are*

$$T\mathcal{M} \simeq (\mathbf{c}, \tilde{\mathbf{c}})_{(1,1)} \quad T\mathcal{K} \simeq (\mathbf{c}, \tilde{\mathbf{a}})_{(1,-1)}. \quad (11.123)$$

(2) *The Zamolodchikov metric on both factor manifolds is special Kähler.*

Proof (1) The splitting is preserved by parallel transport with the Zamolodchikov Levi-Civita connection, which is torsionless; then the statement follows from the Nirenberg–Newlander theorem.

(2) For the manifold $\mathcal{M}$ of a σ -model SCFT it follows from **Claim 11.2** (2). The general case follows from tt^* geometry [4, 5, 13]. $\square$

Note 11.6 The first statement (but not the last one!) holds for all 2d (2, 2) QFT.

Bosonization From the OPE

$$J(z)J(w) \sim \frac{\hat{c}}{(z-w)^2} \quad (11.124)$$

we see that we may bosonize the $U(1)$ current in terms of a free scalar

$$J(z) = i\sqrt{\hat{c}}\partial\phi(z), \quad \phi(z)\phi(w) \sim -\log(z-w), \quad (11.125)$$

so that the (2,2) SCFT decouples in the $c = 1$ free scalar ϕ and a “residual” CFT $\mathcal{T}$ with $c_{\mathcal{T}} = 3\hat{c} - 1$. A SCFT operator $\Phi_{(q,\tilde{q})}(z, \bar{z})$ of conformal charges $(q, \tilde{q})$ and Virasoro weights $(h, \tilde{h})$ must have the form

$$\Phi_{(q,\tilde{q})}(z, \bar{z}) = \exp\left[iq\phi(z)/\sqrt{\hat{c}} + i\tilde{q}\tilde{\phi}(\bar{z})/\sqrt{\hat{c}}\right]O(z, \bar{z}) \quad (11.126)$$

for some operator $O(z, \bar{z})$ in the CFT $\mathcal{T}$ of weights

$$h_O = h - \frac{q^2}{2\hat{c}}, \quad \tilde{h}_O = \tilde{h} - \frac{\tilde{q}^2}{2\hat{c}}. \quad (11.127)$$

Consider the left-moving would-be operator

$$\rho(z) \stackrel{\text{def}}{=} \exp(i\sqrt{\hat{c}}\phi(z)), \quad \text{with } q = \hat{c}, \quad h = \frac{(\sqrt{\hat{c}})^2}{2} \equiv \frac{q}{2}, \quad \tilde{h} = \tilde{q} = 0. \quad (11.128)$$

The chiral²⁸ current $\rho(z)$, *when it exists*, is a chiral primary. We know from the discussion in Sect. 5.1 that $\rho(z)$ is present in a modular invariant (2, 2) SCFT if and only if it is *local* with respect to all operators of the SCFT. In view of the OPE

$$\rho(z)\Phi_{(q,\tilde{q})}(w, \bar{w}) \sim (z-w)^q\Phi_{(q+\hat{c},\tilde{q})}(w, \bar{w}) \quad (11.129)$$

²⁸ “Chiral” has two different meanings in this discussion. The reader should not confuse the two.

this requires $q \in \mathbb{Z}$ for all operators and in particular $\hat{c} \in \mathbb{N}$. For the rest of this chapter we *assume* the following condition:

- ($\ddagger$) for all NS-NS operators q , $\tilde{q} \in \mathbb{Z}$ and $\hat{c} \equiv \tilde{\hat{c}} \in \mathbb{N}$, so that both left- and right-moving operators $\rho(z)$ and $\tilde{\rho}(\bar{z})$ exist.

A chiral primary which saturates the bound $q \leq \hat{c}$ should have the form

$$\rho(z) \cdot (\text{a right-moving operator with } \tilde{h} = \tilde{q}/2). \quad (11.130)$$

Indeed a chiral primary with $q = \hat{c}$ has the form (11.126) with $h_O = 0$. Since the dimension of the chiral ring $(\mathbf{c}, \tilde{\mathbf{c}})$ is finite, say $k + 1$, there is a finite sequence

$$\{1 \equiv O_0, O_1, \dots, O_{k-1}, O_k\} \quad (11.131)$$

of local operators in the $\mathcal{T}$ CFT with Virasoro weights

$$h(O_j) = \frac{q_j}{2} \left(1 - \frac{q_j}{\hat{c}}\right), \quad \tilde{h}(O_j) = \frac{\tilde{q}_j}{2} \left(1 - \frac{\tilde{q}_j}{\hat{c}}\right), \quad (11.132)$$

such that the NS-NS chiral primaries

$$\phi_j(z, \bar{z}) = \exp \left[i \frac{q_j}{\sqrt{\hat{c}}} \phi(z) + i \frac{\tilde{q}_j}{\sqrt{\hat{c}}} \tilde{\phi}(\bar{z}) \right] O_j(z, \bar{z}) \quad (11.133)$$

form a basis of the protected ring $(\mathbf{c}, \tilde{\mathbf{c}})$. Here

$$\{(q_j, \tilde{q}_j)\}_{j=0}^k, \quad \text{with } (q_0, \tilde{q}_0) = (0, 0) \quad (11.134)$$

is the spectrum of $U(1) \times \tilde{U}(1)$ superconformal charges in the ring $(\mathbf{c}, \tilde{\mathbf{c}})$.

Consider now the NS-NS operators

$$\bar{\phi}_j(z, \bar{z}) \stackrel{\text{def}}{=} \rho^\dagger \tilde{\rho}^\dagger C_j(z, \bar{z}) = \exp \left[i \frac{q_j - \hat{c}}{\sqrt{\hat{c}}} \phi(z) + i \frac{\tilde{q}_j - \hat{c}}{\sqrt{\hat{c}}} \tilde{\phi}(\bar{z}) \right] O_j(z, \bar{z}) \quad (11.135)$$

with charges and weights

$$q = q_j - \hat{c} \leq 0, \quad h = \frac{(q_j - \hat{c})^2}{2\hat{c}} + \frac{q_j}{2} \left(1 - \frac{q_j}{\hat{c}}\right) \equiv -\frac{1}{2}(q_j - \hat{c}), \quad (11.136)$$

and similar expressions for $\tilde{q}$ and $\tilde{h}$. The operators $\bar{\phi}_j$ saturate the unitary bound (11.108), hence are *anti*-chiral primaries, i.e. Hermitian conjugates of the chiral primaries. They form a basis of the ring $(\mathbf{a}, \tilde{\mathbf{a}})$. Then we may choose the basis (11.131) so that there is an order 2 permutation ν of $\{0, 1, \dots, k\}$ such that

$$q_j + q_{v(j)} = \hat{c}, \quad O_j(z, \bar{z})^\dagger = O_{v(j)}(z, \bar{z}), \quad \phi_j(z, \bar{z})^\dagger = \bar{\phi}_{v(j)}(z, \bar{z}). \quad (11.137)$$

Likewise, the NS-NS operators

$$\sigma_j(z, \bar{z}) \stackrel{\text{def}}{=} : \tilde{\rho}^\dagger C_j(z, \bar{z}) : = \exp \left[i \frac{q_j}{\sqrt{\hat{c}}} \phi(z) + i \frac{\tilde{q}_j - \hat{c}}{\sqrt{\hat{c}}} \tilde{\phi}(\bar{z}) \right] O_j(z, \bar{z}) \quad (11.138)$$

$$\bar{\sigma}_j(z, \bar{z}) \stackrel{\text{def}}{=} : \rho^\dagger C_j(z, \bar{z}) : = \exp \left[i \frac{q_j - \hat{c}}{\sqrt{\hat{c}}} \phi(z) + i \frac{\tilde{q}_j}{\sqrt{\hat{c}}} \tilde{\phi}(\bar{z}) \right] O_j(z, \bar{z}) \quad (11.139)$$

form a basis of the twisted-chiral rings $(\mathbf{c}, \tilde{\mathbf{a}})$ and $(\mathbf{a}, \tilde{\mathbf{c}})$, respectively. We conclude that, under assumption $(\ddagger)$, the 4 SUSY protected rings are all isomorphic as vector spaces, the isomorphisms being given by multiplication by $\rho(z)$, $\tilde{\rho}(\bar{z})$ and their Hermitian conjugates. This isomorphism yields a Frobenius trace map (**Box 11.2**)

$$\langle \cdots \rangle_{\mathcal{R}} : \mathcal{R} \rightarrow \mathbb{C} \quad (11.140)$$

on the 4 rings $\mathcal{R} = (\mathbf{c}, \tilde{\mathbf{c}})$, $(\mathbf{a}, \tilde{\mathbf{a}})$, $(\mathbf{c}, \tilde{\mathbf{a}})$, $(\mathbf{a}, \tilde{\mathbf{c}})$. For $\mathcal{R} = (\mathbf{c}, \tilde{\mathbf{c}})$ it is given by²⁹

$$\phi(z, \bar{z}) \mapsto \langle \phi \rangle_{(\mathbf{c}, \tilde{\mathbf{c}})} \equiv \lim_{w \rightarrow z} |z - w|^{2\hat{c}} \rho(w)^\dagger \tilde{\rho}(\bar{w})^\dagger \phi(z, \bar{z}) \in \mathbb{C}. \quad (11.141)$$

As before, let $\{\phi_j\}_{j=0}^k$ be a basis of $\mathcal{R}$. We set

$$\eta_{ij} = \langle \phi_i \phi_j \rangle_{\mathcal{R}}. \quad (11.142)$$

η_{ij} is the *topological metric*: it is graded-symmetric and non-degenerate. The tensor

$$C_{ijk} \stackrel{\text{def}}{=} C_{ij}^l \eta_{lk} \equiv \langle \phi_i \phi_j \phi_k \rangle_{\mathcal{R}} \quad (11.143)$$

is totally graded-symmetric in its three indices.

Spin-Fields

Consider the left-moving would-be operator

$$\varepsilon(z) \stackrel{\text{def}}{=} \exp[i\sqrt{\hat{c}}\phi(z)/2], \quad \text{such that } : \varepsilon(z)^2 : \equiv \rho(z), \quad (11.144)$$

$$\varepsilon(z) \Phi_{(q, \bar{q})}(w, \bar{w}) \sim (z - w)^{q/2} \Phi_{q+\hat{c}/2, \bar{q}}(w, \bar{w}). \quad (11.145)$$

The field $\varepsilon(z)$ is local with respect to operators of even $U(1)$ charge $q \in 2\mathbb{Z}$ while it produces a square-root branch-cut in fields of odd charge. Since the supercurrents $G^\pm(z)$ have $q = \pm 1$, $\varepsilon(z)$ introduces square-root cuts in the supercurrents, i.e. $\varepsilon(z)$ is a *spin-field* which creates a Ramond state out of the SL_2 vacuum. Since

²⁹ Note that the RHS of (11.141) is independent of $z, \bar{z}$.

$$h_\varepsilon = \frac{1}{2}(\hat{c}/2)^2 = \frac{1}{8}\hat{c}, \quad (11.146)$$

$\varepsilon(z)$ saturates the unitary bound (11.141) and hence generates a left-moving *Ramond ground state*. More generally, for each $(\mathbf{c}, \tilde{\mathbf{c}})$ primary ϕ_j the operator

$$S_j(z, \bar{z}) \stackrel{\text{def}}{=} \varepsilon^\dagger \phi_j := \exp \left[i \frac{q_j - \hat{c}/2}{\sqrt{\hat{c}}} \phi(z) + i \frac{\tilde{q}_j}{\sqrt{\hat{c}}} \tilde{\phi}(\bar{z}) \right] \mathcal{O}_j(z, \bar{z}) \quad (11.147)$$

is a spin-field mapping NS-NS $\leftrightarrow$ R-NS with charge and Virasoro weight

$$q = q_j - \frac{1}{2}\hat{c}, \quad h \equiv \frac{1}{2\hat{c}}(q_j - \frac{1}{2}\hat{c})^2 + \frac{1}{2}q_j \left(1 - \frac{q_j}{\hat{c}}\right) = \frac{\hat{c}}{8} \quad (11.148)$$

so, acting on the SL_2 -vacuum, S_j generates a state which is a left-moving Ramond ground state and a right-moving NS chiral primary. To shorten the discussion we focus on the left-moving part of the operators; the same story applies to the right side. We conclude that the OPE with $\varepsilon(z)$, $\varepsilon(z)^\dagger$ (resp. $\tilde{\varepsilon}(\bar{z})$, $\tilde{\varepsilon}(\bar{z})^\dagger$) yield linear isomorphisms between the space of left-moving (resp. right-moving) anti-chiral primaries $\mathbf{a}$, Ramond vacua, and chiral primaries $\mathbf{c}$

$$\mathbf{a} \begin{array}{c} \xrightarrow{\varepsilon} \\ \xleftarrow{\varepsilon^\dagger} \end{array} \mathbf{R \text{ vacua}} \begin{array}{c} \xleftarrow{\varepsilon} \\ \xrightarrow{\varepsilon^\dagger} \end{array} \mathbf{c} \quad (11.149)$$

By associativity of the OPE, these maps are in fact isomorphisms of chiral ring modules and anti-chiral ring modules, respectively. These isomorphisms, as well as their compositions, are called the *spectral-flow isomorphisms* [27].

σ -Models with CY Target-Spaces

As we saw in Sect. 11.2, a SUSY σ -model with target space a CY m -fold X is (for a suitable Kähler metric) a $(2, 2)$ SCFT whose conformal manifold is the product $\mathcal{M} \times \mathcal{K}$ with $\dim_{\mathbb{C}} \mathcal{M} = h^{m-1,1}(X)$, $\dim_{\mathbb{C}} \mathcal{K} = h^{1,1}(X)$. We wish to study this SCFT in more detail. The central charge $\hat{c}$, the spectrum of $U(1) \times \tilde{U}(1)$ charges, and the number of protected operators of given charges $(q_j, \tilde{q}_j)$ are constant along the conformal manifold, so they can be computed in any convenient regime, e.g. in the large Kähler class limit, that is, in the classical (weak coupling) limit $\hbar \rightarrow 0$. In other words: *all protected quantities can be safely computed using free field theory*. Then $\hat{c} \equiv m$, the complex dimension, while at extreme weak coupling the (twisted)chiral primaries are written as products of Fermi fields with indices contracted with harmonic tensors on X of the appropriate kind. The left-moving (resp. right-moving) 2d fermions are sections of the appropriate spin-structure line bundle $\mathcal{L}$ (resp. $\tilde{\mathcal{L}}$) twisted by (the pull-back to the world-sheet of) the complexified tangent bundle³⁰ $TX \oplus \bar{TX}$. We use the notation for the Fermi fields in Table 11.1.

³⁰ To simplify the notation, here and below TX (resp. T^*X) stands for the *holomorphic* tangent (resp. cotangent) bundle of vectors (resp. forms) of type $(1, 0)$, $T^*X = (TX)^\vee$.

Table 11.1 σ -model Fermi fields: they are sections of the bundle in the first column. In the last column our (conventional) assignments $(q, \tilde{q})$ for their $U(1) \times \tilde{U}(1)$ charges; the somewhat unnaturally looking choice is for later convenience. Here $(i, \bar{i} = 1, \dots, m)$

Bundle	field	$(q, \tilde{q})$	Bundle	field	$(q, \tilde{q})$
$\mathcal{L} \otimes \phi^* TX$	ψ^i	$(-1, 0)$	$\mathcal{L} \otimes \phi^* T^* X$	$\rho_i = g_{i\bar{j}} \psi^{\bar{j}}$	$(1, 0)$
$\tilde{\mathcal{L}} \otimes \phi^* TX$	$\tilde{\psi}^{\bar{i}}$	$(0, -1)$	$\tilde{\mathcal{L}} \otimes \phi^* T^* X$	$\tilde{\rho}_{\bar{i}} = g_{\bar{i}j} \psi^j$	$(0, 1)$
$\mathcal{L} \otimes \phi^* \bar{T} X$	$\psi^{\bar{i}}$	$(1, 0)$	$\mathcal{L} \otimes \phi^* \bar{T}^* X$	$\rho_{\bar{i}} = g_{\bar{i}j} \psi^j$	$(-1, 0)$
$\tilde{\mathcal{L}} \otimes \phi^* \bar{T} X$	$\tilde{\psi}^{\bar{i}}$	$(0, 1)$	$\tilde{\mathcal{L}} \otimes \phi^* \bar{T}^* X$	$\tilde{\rho}_{\bar{i}} = g_{\bar{i}j} \psi^j$	$(0, -1)$

We consider the following four cohomology rings on X

$$H^\bullet(X, \wedge^\bullet T^* X) = \bigoplus_{q,p} H^q(X, \wedge^p T^* X), \quad H^\bullet(X, \wedge^\bullet TX) = \bigoplus_{q,p} H^q(X, \wedge^p TX), \quad (11.150)$$

$$H^\bullet(X, \wedge^\bullet T^* X)^* = \bigoplus_{p,q} H^p(X, \wedge^q T^* X)^*, \quad H^\bullet(X, \wedge^\bullet TX)^* = \bigoplus_{p,q} H^p(X, \wedge^q TX)^*. \quad (11.151)$$

On a Calabi–Yau m -fold X we have the isomorphisms

$$H^q(X, \wedge^{m-p} TX) \simeq H^q(X, \wedge^p T^* X) \simeq H^p(X, \wedge^q T^* X)^*, \quad (11.152)$$

so the 4 rings have isomorphic underlying vector $\mathbb{C}$ -spaces but are *inequivalent* as Frobenius algebras. Comparing with our discussion of SUSY protected rings, we see that, at extreme weak coupling we have the four $\mathbb{C}$ -algebras isomorphisms³¹

$$\begin{aligned} H^\bullet(X, \wedge^\bullet TX) &\rightarrow (\mathbf{c}, \tilde{\mathbf{c}}), & H^\bullet(X, \wedge^\bullet T^* X) &\rightarrow (\mathbf{c}, \tilde{\mathbf{a}}), \\ H^\bullet(X, \wedge^\bullet TX)^* &\rightarrow (\mathbf{a}, \tilde{\mathbf{a}}), & H^\bullet(X, \wedge^\bullet T^* X)^* &\rightarrow (\mathbf{a}, \tilde{\mathbf{c}}), \end{aligned} \quad (11.153)$$

given by

$$\begin{aligned} \phi^{i_1 \dots i_p} \bar{j}_1 \dots \bar{j}_q \partial_{z^{i_1}} \dots \partial_{z^{i_p}} d\bar{z}^{j_1} \dots d\bar{z}^{j_q} &\mapsto \\ &\mapsto \phi^{i_1 \dots i_p} \bar{j}_1 \dots \bar{j}_q \tilde{\rho}_{i_1} \dots \tilde{\rho}_{i_p} \psi^{\bar{j}_1} \dots \psi^{\bar{j}_q}, \end{aligned} \quad (11.154)$$

$$\begin{aligned} \omega_{i_1 \dots i_p \bar{j}_1 \dots \bar{j}_q} dz^{i_1} \dots dz^{i_p} d\bar{z}^{j_1} \dots d\bar{z}^{j_q} &\mapsto \\ &\mapsto \omega_{i_1 \dots i_p \bar{j}_1 \dots \bar{j}_q} \tilde{\psi}^{i_1} \dots \tilde{\psi}^{i_p} \psi^{\bar{j}_1} \dots \psi^{\bar{j}_q}, \end{aligned} \quad (11.155)$$

and their Hermitian conjugates. In Eqs. (11.154), (11.155) ϕ and ω are *harmonic representatives* of their cohomology classes. From **Table 11.1** we see that the isomorphisms (11.153) identify the charges $(q, \tilde{q})$ with the *type* of the harmonic form³²

³¹ Note that at weak coupling the operators (11.154), (11.155) saturate the unitarity bound $h = |q|/2$ and so are chiral primaries; then they belong to short SUSY protected supermultiplets and their existence, quantum numbers, and dimensions are preserved even at strong coupling. The only aspect which may be renormalized by the world-sheet interactions is their product table.

³² The conventions were chosen to avoid clashes between the two meanings of the symbol q .

$$(q, \tilde{q}) = (q, p) \text{ and, respectively, } (q, \tilde{q}) = (q, -p). \quad (11.156)$$

Equation (11.153) extends to the full chiral rings the isomorphisms (11.123) previously established for their subspaces $|q| = |\tilde{q}| = 1$. Infinitesimal deformation of Kähler class are elements of $H^1(X, T^*X)$ mapped into elements of $(\mathbf{c}, \tilde{\mathbf{a}})$; infinitesimal deformations of complex structure belong to $H^1(X, TX)$ and are mapped into elements of $(\mathbf{c}, \tilde{\mathbf{c}})$. We know from Sect. 11.2.1 that complex moduli are not corrected by quantum effects: in a CY σ -model the quantum chiral ring $(\mathbf{c}, \tilde{\mathbf{c}})$ is given *exactly* by the classical ring $H^*(X, \wedge^\bullet T^*X)$ and the exact Zamolodchikov metric on the $(\mathbf{c}, \tilde{\mathbf{c}})$ factor of the conformal manifold is the WP metric of the complex moduli $\mathcal{M}$.

Let Ω be a parallel $(m, 0)$ -form on the CY m -fold X . Its conjugate $(0, m)$ -form

$$\overline{\Omega} \in H^m(X, \mathcal{O}) \subset H^*(X, \wedge^\bullet T^*X) \cap H^*(X, \wedge^\bullet TX) \quad (11.157)$$

belongs to both rings $(\mathbf{c}, \tilde{\mathbf{c}})$ and $(\mathbf{c}, \tilde{\mathbf{a}})$. It has $(q, \tilde{q}) = (m, 0) \equiv (\hat{c}, 0)$, saturates the unitary bound, and is identified with the unique chiral primary with these charges

$$\overline{\Omega} \rightsquigarrow \rho(z) \equiv: \varepsilon^2(z) \equiv: \exp[i\sqrt{m}\phi(z)] \quad (11.158)$$

that is, to the square of the spectral-flow operator $\varepsilon(z)$. Comparing the geometric isomorphism (11.20) with the spectral flow, we see that the spin-fields $\varepsilon, \varepsilon^\dagger$ should be identified with the two parallel fermions predicted by CY geometry.

★ Relation to Topological Field Theory

As discussed at the beginning of Chap. 2, a Topological Field Theory (TFT) is a QFT whose spacetime symmetry group contains $\mathbf{Diff}^+$. A TFT can be quantized in any oriented smooth manifold M of dimension d and the observables are independent of the metric on M and any other auxiliary geometric structure.³³ We focus on $d = 2$ with M closed. The group $\mathbf{Diff}^+$ is ∞ -transitive, so the correlations of local operators $\langle \phi_1(z_1) \cdots \phi_s(z_s) \rangle$ are independent of the points z_i . In particular, all local observables $\phi_i(z)$ are scalar fields (**Exercise 2.1**). Although a TFT needs not to satisfy *Spin & Statistics*, for simplicity we assume that all local observables are bosonic. If $\{\phi_i(z)\}$ is a basis of the algebra $\mathfrak{A}$ of local operators, the OPE take the form $\phi_i \phi_j = C_{ij}^k \phi_k$ for numerical coefficients C_{ij}^k . The algebra is commutative so $C_{ij}^k = C_{ji}^k$ and has a unity $\phi_0 \equiv 1$, so $C_{0j}^k = \delta_j^k$. We assume $\dim \mathfrak{A} \equiv N + 1 < \infty$. The trace map $\phi \mapsto \langle \phi \rangle_{S^2}$ yields to $\mathfrak{A}$ the structure of a Frobenius algebra; the non-degenerate symmetric matrix $\eta_{ij} = \langle \phi_i \phi_j \rangle_{S^2}$ is the *topological metric*. We choose the basis $\{\phi_i\}$ so that $\langle \phi_i \rangle_{S^2} = \delta_{i,N}$, i.e. $\eta_{0i} = \delta_{i,N}$. Then

$$\langle \phi_{i_1} \cdots \phi_{i_s} \rangle_{S^2} = (C_{i_1} \cdots C_{i_s})_0^N \quad (11.159)$$

where C_i is the $(N + 1) \times (N + 1)$ matrix $(C_i)_j^k \equiv C_{ij}^k$. Since a TFT is in particular a (degenerate) CFT we have a state-operator correspondence and the Hilbert space $\mathbf{H} \simeq \mathfrak{A} \equiv \mathbb{C}^{N+1}$ as a left-module over itself: this is the (topological) spectral-flow isomorphism.

In our case M is a smooth surface of genus g ; since nothing depends on the metric (or the complex structure) we can see M as a large torus T_0 where we cut-off $g - 1$ disks of radius ε ; at the boundary of each disk we glue-in the boundary of a small torus with a disk removed. In the limit $\varepsilon \rightarrow 0$, the effect of gluing-in a small torus, as seen from the large torus T_0 , is the insertion in a point $p \in T_0$ of a local operator $\mathfrak{h} \in \mathfrak{A}$ (the *handle* operator). Then

³³ For an axiomatic approach to TFT see e.g. [74].

$$\langle \phi_1(z_1) \cdots \phi_s(z_s) \rangle_{\text{genus } g} = \langle \phi_1 \cdots \phi_s \mathfrak{H}^{g-1} \rangle_{\text{torus}} = \text{Tr}[C_1 C_2 \cdots C_s H^{g-1}] \quad (11.160)$$

where $\mathfrak{H}\phi_i = H_i^j \phi_j$, and the path integral is computed as the trace over $\mathbf{H}$. Comparing (11.159) with (11.160) for $g = 0$ we get $\mathfrak{H} \equiv \phi_N$, and this determine all correlations. For details see [75].

A QFT is Diff^+ -invariant iff $T_{\mu\nu} = 0$. In a QFT defined *à la* BRST, this requires $T_{\mu\nu}$ to be BRST-exact, $T_{\mu\nu} = \{Q, b_{\mu\nu}\}$. Suppose we have a $(2, 2)$ SCFT theory whose 4 supercharges $Q^\pm$, $\tilde{Q}^\pm$ transform under $U(1)_{\text{spin}} \times U(1)_L \times U(1)_R$ with charges $(\frac{1}{2}, \pm 1)$ and $(-\frac{1}{2}, \pm 1)$, respectively. We redefine the spin operator by mixing it with the $U(1)_{L,R}$ charges $J, \tilde{J}$

$$L_0 - \tilde{L}_0 \rightsquigarrow L_0 - \tilde{L}_0 - \frac{1}{2}(\varepsilon J + \tilde{\varepsilon} \tilde{J}), \quad \varepsilon, \tilde{\varepsilon} = \pm 1. \quad (11.161)$$

In terms of the redefined quantum numbers, for each one of the 4 choices of signs $\varepsilon, \tilde{\varepsilon}$, 2 supercharges have spin zero: we call them $Q, \tilde{Q}$. These odd scalar charges satisfy the algebra

$$Q^2 = \tilde{Q}^2 = Q\tilde{Q} + \tilde{Q}Q = 0, \quad (11.162)$$

so they have the properties required to be BRST charges; cf. Sect. 3.5. The redefinition (11.161) is equivalent to a redefinition of the energy–momentum tensor

$$T_B \rightsquigarrow T'_B \equiv T_B - \frac{1}{2}\varepsilon\partial J, \quad \tilde{T}_B \rightsquigarrow \tilde{T}'_B \equiv \tilde{T}_B - \frac{1}{2}\tilde{\varepsilon}\partial\tilde{J}. \quad (11.163)$$

The replacement of $T_B, \tilde{T}_B$ with $T'_B, \tilde{T}'_B$ is a simple instance of a general strategy called *topological twisting* [76–78] (for a general account see [74]). The modification of the energy–momentum tensor is equivalent to a modification of the coupling to the background metric; pragmatically: one replaces in the original covariant derivatives $\nabla_i, \nabla_{\tilde{j}}$ acting on fields of charges $(q, \tilde{q})$, with $\nabla_i - \frac{1}{2}q\varepsilon\omega_i$, $\nabla_{\tilde{j}} - \frac{1}{2}\tilde{q}\tilde{\varepsilon}\omega_{\tilde{j}}$ where $\omega_i, \omega_{\tilde{j}}$ is the Levi-Civita connection.

Exercise 11.2 Show that T'_B satisfies the Virasoro OPE with $c = 0$.

Acting on operators (resp. states) Q is identified with $G_{-1/2}^\varepsilon$ (resp. G_0^ε). Hence

$$\{Q, G^{-\varepsilon}(z)\} = T'_B(z), \quad (11.164)$$

so that the twisted energy–momentum tensor T'_B is BRST-exact, and the twisted model is a TFT. The physical local operators are the BRST-closed ones modulo the BRST-exact ones. Depending on the choice of signs $\varepsilon, \tilde{\varepsilon}$ the ring $\mathcal{R}_{\text{TFT}}$ of topological operators gets identified with one of the 4 SUSY protected rings: $\varepsilon = +1 \leftrightarrow \mathbf{c}, \varepsilon = -1 \leftrightarrow \mathbf{a}$, and the same rule for the tilded symbols.

A TFT has no singularities at short or long distances, simply because distance has no physical meaning in a Diff^+ -invariant theory. In particular there cannot be *any* renormalization.

In flat space the modification of the covariant derivative is trivial, since $\omega_i = 0$, and hence the twisted and original theories have the same action and path integral measure. The only difference is that in the twisted theory we *declare* that the only allowed insertions are the operators in the protected ring $\mathcal{R}_{\text{TFT}}$. Hence, in flat space, correlation functions of protected operators in a chiral ring may be seen both as SCFT or TFT correlators. In the second interpretation they are trivially non-renormalized; from the SCFT viewpoint this fact looks as a non-trivial non-renormalization theorem. Most SUSY non-renormalization theorems arise from topological twists.

There are two basic twists: either $\tilde{\varepsilon} = -\varepsilon$ or $\tilde{\varepsilon} = \varepsilon$. One says [77] that the first TFT is the *A*-model, and the second one the *B*-model (compare with the dichotomy Type IIA/IIB).

Note 11.7 Previously we claimed without proof that the OPE coefficients C_{ij}^k of the $(\mathbf{c}, \tilde{\mathbf{c}})$ ring depend holomorphically on $\mathcal{M}$. Since $C_{ij}^k \langle \phi_k \phi_l \rangle_{\text{TFT}} = \langle \phi_i \phi_k \phi_l \rangle_{\text{TFT}}$, it is enough to show that the TFT correlation functions are holomorphic. $\partial_{s^a} \in T\mathcal{M}$ correspond to an F-term deformation

$\int d^2\theta \phi_a$ with $\phi_a \in (\mathbf{c}, \tilde{\mathbf{c}})$, while $\partial_{\bar{s}a} \in \overline{TM}$ to a deformation $\int d^2\bar{\theta} \bar{\phi}_a$ which, from the TFT view-point, is BRST-exact $\int d^2\bar{\theta} \bar{\phi}_a \equiv \{Q, \bar{\eta}_a\}$ with $\bar{\eta}_a \equiv [\tilde{Q}, \bar{\phi}_a]$. BRST invariance yields

$$\partial_{\bar{s}a} \langle \phi_{i_1} \cdots \phi_{i_s} \rangle_{\text{TFT}} = \int d^2z \{Q, \bar{\eta}_a(z)\} \phi_{i_1} \cdots \phi_{i_s} \rangle_{\text{TFT}} = 0. \quad (11.165)$$

★ $\mathcal{N} = 2$ Minimal Models vs. $\widehat{\mathfrak{su}(2)}$ Kač-Moody CFT

The bosonization of the conformal $U(1)$, Eqs. (11.124)–(11.149), yields for the supercurrents

$$G^\pm(z) = \exp\left[\pm i\phi(z)/\sqrt{\hat{c}}\right] P^\pm(z) \quad (11.166)$$

for some chiral currents $P^\pm(z)$ of the residual CFT $\mathcal{T}$ with weight

$$h_P = \frac{3\hat{c} - 1}{2\hat{c}} \equiv 1 + \frac{\hat{c} - 1}{2\hat{c}} \geq 0. \quad (11.167)$$

Consider the case that $h_P < 1$, i.e. $\hat{c} < 1$. The free superfield has $\hat{c} = 1$, so the condition $\hat{c} < 1$ for a $\mathcal{N} = 2$ SCFT is analogue to $c < 1$ for Virasoro which is consistent with unitarity only for a discrete set of values c associated to the Virasoro minimal models [79, 80]. The $\mathcal{N} = 2$ SCFT with $\hat{c} < 1$ are called $\mathcal{N} = 2$ *minimal models* and again unitarity restricts $\hat{c}$ to a discrete set [81–83]. Consider the chiral currents

$$J^\pm(z) = \exp\left[\pm i\sqrt{\frac{1-\hat{c}}{\hat{c}}}\phi(z)\right] P^\pm(z), \quad J^0(z) = i\sqrt{\frac{\hat{c}}{1-\hat{c}}}\partial\phi(z) \quad (11.168)$$

which, by construction, have weights $(h, \tilde{h}) = (1, 0)$, so are conserved currents generating a Kač-Moody (affine) Lie algebra; cf. **Box 7.1**. Since

$$J^0(z) J^\pm(w) \sim \pm \frac{J^\pm(w)}{z-w}, \quad J^0(z) J^0(w) \sim \frac{\hat{c}}{1-\hat{c}} \frac{1}{(z-w)^2}, \quad (11.169)$$

we conclude that these currents generate the $SU(2)$ Kač-Moody Lie algebra at level

$$\frac{k}{2} = \frac{\hat{c}}{1-\hat{c}} \Rightarrow c \equiv 3\hat{c} = \frac{3k}{k+2} \equiv c_{\text{Sugawara}} \quad (11.170)$$

so that the Sugawara central charge saturates the central charge of the free scalar plus the $\mathcal{T}$ CFT. Thus the minimal $\mathcal{N} = 2$ SCFT and the $SU(2)_k$ current algebra are just two different maximal local projections of the same underlying CFT. The $SU(2)_k$ CFTs are classified by the *ADE* simply-laced Lie algebras [80] (cf. **Appendix 2** of Chap. 2). In virtue of the above correspondence the minimal $\mathcal{N} = 2$ are also classified by the *ADE* Lie algebras.

Exercise 11.3 Describe the SCFT chiral ring for each simply-laced Lie algebra $\in ADE$.

★ Gepner Models

We are interested in $(2, 2)$ SCFTs which satisfy (‡). One way to construct such a SCFT is to consider a s -tuple of non-interacting minimal models $\mathcal{T}_1, \dots, \mathcal{T}_s$ with $\sum_i \hat{c}_i \in \mathbb{N}$. Bosonize the $U(1)$ current of $\mathcal{T}_i$ with the scalar ϕ_i and consider the operator $\tilde{\Omega} \equiv \exp(i \sum_i \sqrt{c_i} \phi_i)$. The maximal local subalgebra containing $\tilde{\Omega}$ is a SCFT which satisfies condition (‡): typically it turns out to be a CY σ -model at a special point in the conformal manifold $\mathcal{C}$. Writing it as a tensor product of minimal models (whose correlators are exactly known by their relation with the Kač-Moody algebras) gives a simple way of solving non-perturbatively the CY σ -model at these special points. These points are called *Gepner points* and the corresponding $(2, 2)$ SCFT *Gepner models* [84–86].

11.3 (2,2) SCFTs as Type II Backgrounds

Let $\mathcal{T}$ be a (2, 2) SCFT with $\hat{c} = m \leq 4$ which satisfies condition $(\ddagger)$. We can use $\mathcal{T}$ plus $d = 10 - 2m$ free (1, 1) superfields as the world-sheet theory of either Type IIA or Type IIB. This gives a string model with Poincaré invariance in d dimensions, i.e. a compactification of the superstring from 10d to $(10 - 2m)d$. The gauge fixed world-sheet action takes the form ($\mu = 0, 1, \dots, 9 - 2m$)

$$S = S_{\mathcal{T}} + \int d^2z \left(\partial X_{\mu} \bar{\partial} X^{\mu} + \psi_{\mu} \bar{\partial} \psi^{\mu} + \tilde{\psi}_{\mu} \partial \tilde{\psi}^{\mu} \right) + \text{ghosts}. \quad (11.171)$$

GSO Projection for Type IIA/Type IIB

We bosonize the free fermions ψ^{μ} with the scalars φ_a ($a = 1, \dots, 5 - m$), the β, γ ghosts with the free scalar φ (see Chaps. 2, 3), and the SCFT $U(1)$ current with the free scalar ϕ ; cf. Eq. (11.125). An operator takes the form

$$\exp \left[p \varphi + i \sum_a s_a \varphi_a + i q \phi / \sqrt{\hat{c}} \right] \exp \left[\tilde{p} \tilde{\varphi} + i \sum_a \tilde{s}_a \tilde{\varphi}_a + i \tilde{q} \tilde{\phi} / \sqrt{\hat{c}} \right] \mathcal{U}, \quad (11.172)$$

where p is its picture charge, s_a its $Spin(9 - 2m, 1)$ weights, q its SCFT $U(1)$ charge, and $\mathcal{U}$ an operator with zero charges; on the right side we have the same story. Comparing with the 10d superstring ($\mathcal{T}$ free), we see that the charges p, s_a, q are *all integral* in the NS sector and *all half-integral* in the R sector. The GSO $\pm$ projections take the form (for the left side: same story on the right)

$$p + \sum_a s_a + q = \begin{cases} 0 \bmod 2 & \text{in NS} \\ \frac{1}{2}(1 \mp 1) \bmod 2 & \text{in R.} \end{cases} \quad (11.173)$$

Massless Physical Vertices and SUSY

To make the story short, we consider only the physically most relevant case $\hat{c} = 3$, i.e. a compactification down to 4d.

Supersymmetry First we look for BRST-invariant vertices of 4d massless gravitini whose existence implies local (unbroken) SUSY. They belong to the NS-R and R-NS sectors. In (say) picture $(-1, -\frac{1}{2})$ the NS-R gravitino vertices read³⁴

$$e^{-\varphi(z) - \tilde{\varphi}(\bar{z})/2} \psi^{\mu}(z) \exp \left[\pm \frac{i}{2} \tilde{\varphi}_1 \pm \frac{i}{2} \tilde{\varphi}_2 \right] \tilde{U}(\bar{z}) e^{ikX}, \quad (11.174)$$

where $\tilde{U}(\bar{z})$ is an operator of the (2, 2) SCFT $\mathcal{T}$ with

$$q = 0 \bmod 2, \quad \tilde{q} = \pm \frac{3}{2} \bmod 1, \quad h = (0, \frac{3}{8}). \quad (11.175)$$

³⁴ The gravitino vertices are the γ -traceless part of (11.174); the γ -traces are vertices for dilatini.

In a non-degenerate (2, 2) SCFT there are only 2 such $\tilde{U}$: the right-side spectral flows

$$\tilde{\varepsilon}(\bar{z})^{\pm} \equiv \exp[\pm i\sqrt{3}\tilde{\phi}(\bar{z})/2]. \quad (11.176)$$

The $\text{GSO}\pm$ projection then keeps the vertices

$$\begin{aligned} e^{-\varphi(z)-\tilde{\varphi}(\bar{z})/2} \psi^{\mu}(z) \tilde{S}_{\alpha}(\bar{z}) e^{i\sqrt{3}\tilde{\phi}(\bar{z})/2} e^{ikX} & (\gamma_5 \tilde{S})_{\alpha} = \pm \tilde{S}_{\alpha} \\ e^{-\varphi(z)-\tilde{\varphi}(\bar{z})/2} \psi^{\mu}(z) \tilde{S}_{\alpha}(\bar{z}) e^{-i\sqrt{3}\tilde{\phi}(\bar{z})/2} e^{ikX} & (\gamma_5 \tilde{S})_{\alpha} = \mp \tilde{S}_{\alpha}, \end{aligned} \quad (11.177)$$

where $\tilde{S}_{\alpha}(\bar{z})$ are right-moving $SO(3, 1)$ spin-fields. Equation (11.177) yields one 4d Majorana gravitino from the NS-R sector plus another one from the R-NS sector. The eight spacetime supercharges (for, say, IIB) are given by the contour integrals of the (anti-)holomorphic world-sheet conserved currents (here $\alpha, \dot{\alpha} = 1, 2$)

$$\begin{aligned} e^{-\tilde{\varphi}(\bar{z})/2} \tilde{S}_{\alpha}(\bar{z}) e^{i\sqrt{3}\tilde{\phi}(\bar{z})/2}, & \quad e^{-\tilde{\varphi}(\bar{z})/2} \tilde{S}_{\dot{\alpha}}(\bar{z}) e^{-i\sqrt{3}\tilde{\phi}(\bar{z})/2}, \\ e^{-\varphi(z)/2} S_{\alpha}(z) e^{i\sqrt{3}\phi(z)/2}, & \quad e^{-\varphi(z)/2} S_{\dot{\alpha}}(z) e^{-i\sqrt{3}\phi(z)/2}, \end{aligned} \quad (11.178)$$

whose internal parts are the operators $e^{\pm i\sqrt{3}\phi/2}, e^{\pm i\sqrt{3}\tilde{\phi}/2}$ which yield the spectral-flow isomorphisms between the NS and R sectors.

The $\text{GSO}+$ projections (11.173) on both sides of IIB guarantee that the 2d SUSY currents (11.178) are local to all physical operators. The SUSY currents are global closed 1-forms on Σ for arbitrary insertions. The 4d SUSY Ward identities follow from contour manipulations as in Chap. 3. The 2d spectral-flow isomorphism becomes a spacetime symmetry interchanging boson $\leftrightarrow$ fermions, i.e. spacetime SUSY. Thus:

To have a 4d supersymmetric Poincaré-invariant vacuum, the internal SCFT must contain a $\mathcal{N} = 2$ SCFT algebra with NS operators satisfying (‡) and we must impose independent $\text{GSO}\pm$ projections on the left- and right-movers

Massless Spectrum: the Universal Sector We return to the Type II massless vertices in 4d. In the NS-NS sector with picture $(-1, -1)$ we have

$$e^{-\varphi-\tilde{\varphi}} \psi_{\mu} \tilde{\psi}_{\nu} e^{ikX}, \quad \mu, \nu = 0, \dots, 3 \quad (11.179)$$

which yields the 4d metric $g_{\mu\nu}$, 2-form gauge field $B_{\mu\nu}$, and 4d dilaton³⁵ Φ . In 4d a 2-form gauge field B is dual to a Peccei–Quinn scalar σ :

$$d\sigma = *dB. \quad (11.180)$$

³⁵ Not to be confused with the 10d one.

The SUSY partners of these “universal” 4d bosonic fields ($g_{\mu\nu}$, σ , and Φ) are 2 Weyl gravitini and 2 Weyl dilatini with vertices (11.174) and the RR bosons with vertices

$$e^{-(\varphi+\tilde{\varphi})/2} (S^t C \gamma^{\mu_1 \cdots \mu_s} \tilde{S}) \varepsilon^{\pm 1} \tilde{\varepsilon}^{\pm 1} e^{ikX}. \quad (11.181)$$

Taking into account the GSO projections and the BRST condition, (11.181) yield one gauge vector A_μ and two 0-form gauge field (“axions”) ξ_μ . In total this “universal” sector, common to all compactifications on a 2d (2,2) SCFT, consists of the graviton, 2 gravitini, a vector, 4 scalars and 2 spin-1/2 fermions. This result holds for the universal sector in both Type IIA and IIB. The interchange IIA $\leftrightarrow$ IIB, i.e. the flip of GSO sign on the right side, is equivalent to flipping $\tilde{\phi} \leftrightarrow -\tilde{\phi}$ in all vertices.

The Full 4d Massless Spectrum The massless spectrum in 4d may be obtained by acting on the NS-NS vertices with the spectral flows, i.e. with 4d SUSY. The non-universal massless NS-NS vertices describe spacetime scalars. In the picture -1 , the GSO projection (11.173) for the non-universal massless NS vertices reduces to $-1 + q = 0 \bmod 2$, so $|q| \geq 1$. By 2d unitarity

$$h \geq \frac{1}{2}|q| \geq \frac{1}{2}, \quad (11.182)$$

and the vertex is massless iff all inequalities are saturated, which in particular entails $h = |q|/2$. We conclude that the internal SCFT part of a massless non-universal NS-NS vertex belongs to one of the four protected rings $(c, \tilde{c})$, $(c, \tilde{a})$, $(a, \tilde{c})$, and $(a, \tilde{a})$; their conformal charges are $(q, \tilde{q}) = (\pm 1, \pm 1)$.

When the SCFT is a σ -model with target a 3-CY manifold X , such operators are elements of $H^1(X, T^*X)$, $H^1(X, TX)$ and their complex conjugates.³⁶ This yields

$$2(h^{1,1} + h^{2,1}) \quad (11.183)$$

4d massless scalars from the NS-NS sector. Interchanging IIB $\leftrightarrow$ IIA flips the sign of $\tilde{\phi}$ in the vertices which has the effect

$$(c, \tilde{c}) \leftrightarrow (c, \tilde{a}) \quad (a, \tilde{c}) \leftrightarrow (a, \tilde{a}). \quad (11.184)$$

A harmonic $(1, 1)$ -form ω on X yields a $(c, \tilde{a})$ primary of the form

$$e^{i(\phi-\tilde{\phi})/\sqrt{3}} \mathcal{U}_\omega, \quad (11.185)$$

with $(q, \tilde{q}) = (1, -1)$. This operator (times e^{ikX}) is a physical massless NS-NS scalar vertex in picture $(0, 0)$. Together with its complex conjugate, it gives 2 scalars per harmonic $(1, 1)$ form on X . The spectral flow maps the NS-NS operator (11.185) into a R-R vacuum with $(q, \tilde{q}) = (-1/2, 1/2)$.

³⁶ Although for $H^1(X, T^*X)$ the identification is exact only at weak coupling, the counting of states is exact for all couplings since these states form short SUSY representations.

In Type IIB Eq. (11.173) reduces to

$$s_1 + s_2 = 0 \bmod 2, \quad \tilde{s}_1 + \tilde{s}_2 = 1 \bmod 2, \quad (11.186)$$

so the R-R vertex associated to the NS-NS vertex (11.185) is

$$e^{-(\varphi+\tilde{\varphi})/2} S_\alpha (\gamma_\mu)^{\alpha\dot{\beta}} \tilde{S}_{\dot{\beta}} e^{-i(\phi-\tilde{\phi})/(2\sqrt{3})} \mathcal{U}_\omega e^{ikX} \quad (11.187)$$

which is the vertex of a scalar “axion” Q (it couples only through $\partial_\mu Q$). Together with its conjugate, it gives 2 additional axion scalars per harmonic (1,1)-form, i.e. per infinitesimal deformation of the Kähler moduli.

A harmonic Kodaira–Spencer vector ϕ on X yields a $(\mathbf{c}, \tilde{\mathbf{c}})$ primary of the form

$$e^{i(\varphi+\tilde{\varphi})/\sqrt{3}} \mathcal{U}_\phi \quad \text{with } (q, \tilde{q}) = (1, 1) \quad (11.188)$$

which, together with its complex conjugate, yields 2 massless scalars per infinitesimal deformation of the complex structure. Spectral flow then gives a corresponding R-R vacuum with $(q, \tilde{q}) = (-1/2, -1/2)$. In Type IIB the GSO projection requires

$$s_1 + s_2 = 0 \bmod 2, \quad \tilde{s}_1 + \tilde{s}_2 = 0 \bmod 2, \quad (11.189)$$

so the surviving R-R vertex

$$e^{-(\varphi+\tilde{\varphi})/2} S_\alpha (\gamma_{\mu\nu})^{\alpha\beta} \tilde{S}_\beta e^{-i(\phi+\tilde{\phi})/(2\sqrt{3})} \mathcal{U}_\phi e^{ikX} \quad (11.190)$$

couples to the self-dual part of a (complex) gauge-vector field strength. Together with its conjugate, this gives one gauge vector per harmonic KS vector, i.e. per infinitesimal deformation of the complex structure of X :

$$\text{Interchanging IIB} \leftrightarrow \text{IIA flip the sign } \tilde{\phi} \leftrightarrow -\tilde{\phi}, \text{ so } (\mathbf{c}, \tilde{\mathbf{c}}) \leftrightarrow (\mathbf{c}, \tilde{\mathbf{a}}) \text{ i.e.}$$

$$\text{IIB} \leftrightarrow \text{IIA} \Rightarrow (\text{complex moduli}) \leftrightarrow (\text{Kähler moduli})$$

$$\Rightarrow \mathcal{M} \leftrightarrow \mathcal{K}.$$

(11.191)

Summary of Massless Spectrum The spacetime 4d massless bosonic d.o.f. for a Type II superstring compactified on a strict 3-CY X of Hodge numbers $h^{1,1}, h^{2,1}$ are

$$\begin{aligned} \text{Type IIA: } & 4\text{d metric, } h^{1,1} + 1 \text{ vectors, } 4h^{2,1} + 2h^{1,1} + 4 \text{ scalars;} \\ \text{Type IIB: } & 4\text{d metric, } h^{2,1} + 1 \text{ vectors, } 4h^{1,1} + 2h^{2,1} + 4 \text{ scalars.} \end{aligned} \quad (11.192)$$

The abstract SUSY theorem about the factorization of the conformal manifold in two spaces—with the first one not corrected by quantum effects (**Claim 11.2**)—remains true, after the GSO projection, for the theory (11.171) which then has a moduli space of SUSY vacua $\mathcal{M} \times \mathcal{Q}$, with $\mathcal{M}$ the complex moduli with its

canonical Weil–Petersson metric and Q a complex space of dimension $4(h^{1,1} + 1)$ which contains $\mathcal{K}$ as a totally geodesic submanifold but has other directions spanned by the scalars Φ , σ , and the RR axions. See Sect. 11.9.

11.4 Mirror Symmetry

An abstract $(2, 2)$ SCFT with $\hat{c} = 3$ and integral NS-NS $U(1)$ charges looks very similar to a SUSY σ -model with target space a 3-CY. One may wonder whether the latter class exhausts all $(2, 2)$ SCFTs. To address the question, we argue as follows: an abstract $(2, 2)$ SCFT has 2 kinds of marginal deformations, the operators in $(\mathbf{c}, \tilde{\mathbf{c}})_{(1,1)}$ and those in $(\mathbf{c}, \tilde{\mathbf{a}})_{(1,-1)}$, and its conformal manifold is a product $\mathcal{M} \times \mathcal{K}$. Suppose that the space $\mathcal{K}$ has positive dimension and there is some region at infinity in $\mathcal{K}$ where $\mathcal{K}$ looks asymptotically as the large-volume region of a Kähler cone. Such a limit is called a MUM point.³⁷ In this limit our “abstract” SCFT would look like a weakly coupled σ -model with a large-volume target space. Since the geometry of the conformal manifold of $(2, 2)$ SCFTs is pretty well understood, one can show that “most” $(2, 2)$ SCFT are indeed CY σ -models. We may think of the class of all $(2, 2)$ SCFT as a “slight” generalization of such σ -models.

In $(2, 2)$ SCFT the distinction between the two rings $(\mathbf{c}, \tilde{\mathbf{c}})$ and $(\mathbf{c}, \tilde{\mathbf{a}})$ is purely conventional: the field redefinition $\tilde{\varphi} \leftrightarrow -\tilde{\varphi}$ interchanges the two rings as well as the two factor manifolds $\mathcal{K} \leftrightarrow \mathcal{M}$, so for the same $(2, 2)$ SCFT we may define a second σ -model using a MUM point of the complex moduli space $\mathcal{M}$ (if present): each such point yields a weakly-coupled σ -model which underlies the same abstract SCFT. This leads to the following conjecture, for which there is an overwhelming evidence. Under certain “mild” assumptions it is a mathematical theorem [70, 74, 89, 90]:

Conjecture 11.2 For each CY m -fold X there is a **mirror** CY m -fold $\check{X}$ such that

$$\begin{aligned} h^{1,1}(X) &= h^{m-1,1}(\check{X}), & h^{m-1,1}(X) &= h^{1,1}(\check{X}) \\ \mathcal{M}(X) &= \mathcal{K}(\check{X}) & \mathcal{K}(X) &= \mathcal{M}(\check{X}). \end{aligned} \tag{11.193}$$

Compactifying IIA on X is equivalent to compactifying IIB on $\check{X}$ and viceversa

So stated, *mirror symmetry* cannot be exactly right for a trivial reason. As mentioned before, between the several thousands known 3-CY, there are a few *rigid* ones without complex deformations; they have $h^{2,1}(X) = 0$, so, applying (11.193), we would get $h^{1,1}(\check{X}) = 0$ which is impossible for a geometrical Calabi–Yau variety since it has at least one harmonic $(1, 1)$ -form, i.e. the Kähler form. However, for the vast majority of the known CYs a mirror exists. In fact even the rigid Calabi–Yau have mirrors but they are “non-geometric” Calabi–Yau. By definition, a “non-geometric” Calabi–Yau of dimension m is just a $(2, 2)$ SCFT satisfying (§) with $\hat{c} = m$ which has

³⁷ MUM=maximally unipotent monodromy [87, 88].

no weakly coupled point at infinity in its $\mathcal{K}$ space. The space $\mathcal{K}$ reduces to a point for the mirror of a rigid CY, so it fits in the definition of “non-geometric” Calabi–Yau. At the (cheap) price of slightly enlarging the notion of “Calabi–Yau σ -model”, mirror symmetry gets restored. It is yet another example of a *stringy duality*, and one of the best established at the math level [70, 74, 90].

Now we arrive at the payoff of mirror symmetry [91]. We know that complex moduli space $\mathcal{M}(X)$ does not get corrected by quantum effects, then, provided we know the mirror $\check{X}$ of our Calabi–Yau X , we can compute the *exact* non-perturbative conformal manifold of its σ -model, with its exact Zamolodchikov metric, as

$$\mathcal{M}(X) \times \mathcal{M}(\check{X}). \quad (11.194)$$

In facts, there are several other physically interesting quantities that we may compute exactly from the datum of a *mirror pair* $X, \check{X}$ of Calabi–Yau. The most spectacular application of mirror symmetry is to the *enumerative geometry* of algebraic varieties [91–94]. It is a wonderful topic, but quite afiel of the present elementary introduction to string theory.

11.5 Heterotic $E_8 \times E_8$ on a Calabi–Yau 3-fold

Calabi–Yau manifolds may be also used to compactify the heterotic string from 10d to 4d while preserving $\mathcal{N} = 1$ supersymmetry.³⁸ The 2d action for the compactified theory is the same as before, Eq. (11.171), except that now we do not have left-moving ψ^μ nor β, γ ghosts; to cancel the Weyl anomaly on the left side we need to add a CFT with central charges $(c, \bar{c}) = (13, 0)$ which we realize as 26 left-moving fermions

$$\lambda(z)^a, \quad a = 1, \dots, 16 \quad \text{and} \quad \chi(z)^m, \quad m = 1, \dots, 10. \quad (11.195)$$

In the $E_8 \times E_8$ heterotic string, the λ^a generate the current algebra of the first E_8 gauge group, and we enforce the GSO projection on this set of 16 fermions as in Chap. 7. The χ^m ’s may be seen are a subset of the fermions in the second block of 16 left-moving fermions which in the original 10d theory generate the second gauge factor E_8 (cf. Chap. 7). Hence the $\chi^m \chi^n$ are the chiral currents of a subgroup $SO(10)$ of the second E_8 . The original 10d GSO projection in the second block of left-moving fermions becomes a condition relating the left-moving conformal $U(1)$ charge q of the internal $(2, 2)$ SCFT to the $SO(10)$ weights of the physical vertex

$$\exp \left[i \sum_a s_a \phi_a + i q \varphi / \sqrt{3} \right] \mathcal{U}. \quad (11.196)$$

³⁸ In the heterotic case only the 4 right-moving supercharges are present.

The s_a and q are either all integral or all half-integral. The GSO projection requires

$$\sum_a s_a + q \in 2\mathbb{Z}. \quad (11.197)$$

The spacetime gauge symmetries are given by the 2d left-moving currents, i.e. the Virasoro primaries with $(h, \tilde{h}) = (1, 0)$ which survive the GSO projections. They are the world-sheet currents of the first E_8 which are the same ones as in 10d, and

$$\chi^m \chi^n, \quad S_\alpha e^{i\sqrt{3}\phi/2}, \quad S_{\dot{\alpha}} e^{-i\sqrt{3}\phi/2}, \quad i\partial\varphi, \quad (11.198)$$

where S_α (resp. $S_{\dot{\alpha}}$) are the chirality +1 (resp. -1) spin-fields of the $SO(10)$ current algebra of dimension $\frac{5}{8}$ generated by the χ^m . The currents transform in the following representations of the manifest $SO(10)$ symmetry:

$$\mathbf{45} \oplus \mathbf{16} \oplus \overline{\mathbf{16}} \oplus \mathbf{1} = \mathbf{78} \quad \text{the adjoint } \mathfrak{e}_6 \text{ of } E_6, \quad (11.199)$$

so the 4d gauge group is $E_8 \times E_6$, the centralizer of $SU(3)$ in the 10d group. This group has no cubic **ad**-invariant, so the 4d pure gauge anomaly polynomial vanishes.

The fact that the Type II superstring compactified on this (2,2) SCFT was modular invariant implies that the heterotic theory is also modular invariant, hence anomaly-free. We stress that this construction automatically implements the stringy constraint

$$[\text{tr}(R \wedge R)] = [\text{tr}(F \wedge F)]. \quad (11.200)$$

Indeed 6 out of the original 32 free fermions λ^A now are identified with the left-moving fermions of the $(2, 2)$ σ -model coupled to the CY metric and curvature. The term in the 2d σ -model action (11.84), which in Type II was a 4-fermion coupling with coefficient the spacetime Riemannian curvature of the internal manifold

$$R_{i\bar{j}k\bar{l}} \tilde{\psi}^i \tilde{\psi}^{\bar{j}} \psi^k \psi^{\bar{l}} \rightsquigarrow F_{i\bar{j}k\bar{l}} \tilde{\psi}^i \tilde{\psi}^{\bar{j}} \lambda^k \lambda^{\bar{l}}, \quad (11.201)$$

in the heterotic string gets interpreted as a coupling involving the gauge curvature 2-form $F_{i\bar{j}}$. In other words: the gauge bundle over the 10d space $\mathbb{R}^{3,1} \times X$ is a trivial $E_6 \times E_8$ bundle (from the 2d free fermions) times a $SU(3)$ bundle $\equiv$ holomorphic tangent bundle of the 3-CY X . The background gauge connection A_i is set equal to the spin-connection ω_i on X , and hence $R \equiv F$ by construction, and the constraint (11.200) is trivially satisfied.

The adjoint of E_8 decomposes into representations of $SU(3) \times E_6 \subset E_8$ as

$$\mathfrak{e}_8 = \mathfrak{su}(3) \oplus \mathfrak{e}_6 \oplus (\mathbf{3}, \overline{\mathbf{27}}) \oplus (\overline{\mathbf{3}}, \mathbf{27}). \quad (11.202)$$

Because of the identification of the gauge and spin-connection, the components of the internal E_8 gauge field which are charged under the unbroken E_6 have the form³⁹

$$\begin{aligned} \mathbf{27} &: A_{[ij],\bar{k}}^a & A_{[ij],k}^a, \\ \overline{\mathbf{27}} &: A_{aj,\bar{k}} & A_{aj,k}. \end{aligned} \quad (11.203)$$

It is enough to consider the fields in the second column since the other ones are their complex conjugates. Expanding them in harmonic forms on X , we see that we have $h^{2,1}$ complex massless scalars in the $\mathbf{27}$ and $h^{1,1}$ in the $\overline{\mathbf{27}}$.

Exercise 11.4 Show that the massless spectrum of the $E_8 \times E_8$ heterotic string compactified on a Calabi–Yau 3-fold consists of (i) the 4d $\mathcal{N} = 1$ gravity supermultiplet: the graviton and a Majorana gravitino; (ii) a chiral superfield containing a dilaton, an axion and the dilatino; (iii) gauge bosons in the adjoint of $E_6 \times E_8$; (iv) $h^{2,1}$ chiral superfields in the $\mathbf{27}$ of E_6 ; (v) $h^{1,1}$ chiral superfields in the $\overline{\mathbf{27}}$ of E_6 ; (vi) $h^{2,1}$ chiral superfields for the complex moduli; (vii) $h^{1,1}$ chiral superfields for the Kähler moduli; (viii) $\dim H^1(\text{End } TX)$ chiral superfields neutral for E_6 .

In particular, seeing the low-energy 4d theory as a “standard” E_6 GUT [95]

$$\frac{\text{net number of generations}}{\#(\text{of } \mathbf{27}) - \#(\text{of } \overline{\mathbf{27}})} = h^{2,1} - h^{1,1} = -\frac{1}{2} \chi(X). \quad (11.204)$$

$SO(32)$ Heterotic String on a 3-CY As before, the 4d gauge group G_4 is the centralizer of $SU(3)$ in the 10d gauge group G . For $G = SO(32)$ this gives $G_4 = U(1) \times SO(26)$, whose anomaly polynomial is $\frac{1}{2}\chi(X)F^3$ with F the $U(1)$ field strength. For most CYs this polynomial is non-zero. However the theory is modular invariant and by **Theorem 9.1** the 4d gauge anomaly is canceled by a one-loop Green–Schwarz counter-term (plus a 2-loops seagull)

$$\propto \int B \wedge F \rightsquigarrow \propto \int d^4x A^\mu \partial_\mu \sigma \quad (11.205)$$

σ is a Goldstone boson which is eaten by the $U(1)$ gauge vector *à la* Higgs. The vector becomes massive. It may trigger spontaneous SUSY breaking too; cf. [96].

³⁹ **Index convention:** upper (lower) index $a = 1, 2, \dots, 27$ index of $\mathbf{27}$ ($\overline{\mathbf{27}}$) of E_6 . Lower indices $i, j = 1, 2, 3$ fundamental index of $SU(3)$. An index of the anti-fundamental of $SU(3)$ is written as an antisymmetric pair of fundamental indices $[ij]$. The indices before (resp. after) the comma originate from the gauge (resp. vector) indices of the 10d gauge vector field.

11.6 Type II Compactified on a 3-CY: the Spacetime Perspective

We consider Type IIA/IIB moving in a 10d bosonic background Poincaré invariant in $d < 10$ dimensions. The spacetime has the form $M \equiv \mathbb{R}^{d-1,1} \times X_{10-d}$ with X_{10-d} a compact Riemannian manifold of dimension $10 - d$, so that the physics at length scales $\gg$ the size of X_{10-d} is effectively d -dimensional. We are interested in backgrounds which preserve *some* of the original 32 supercharges in 10d since this guarantees stability of the solutions and analytic control on protected quantities.

Conditions for Supersymmetry

A background preserves the supersymmetry of Grassmann parameter ϵ iff the corresponding SUSY variations vanish for *all* background fields Ψ

$$\delta_\epsilon \Psi = 0. \quad (11.206)$$

Our backgrounds are purely bosonic, so the SUSY variation of all bosonic fields vanishes. We must impose that the variations of the fermionic field vanish as well

$$\epsilon\text{-invariant bosonic background} \Leftrightarrow \delta_\epsilon(\text{fermionic fields}) = 0. \quad (11.207)$$

In (say) 10d Type IIB, the fermions are a Weyl gravitino ψ_M and a Weyl dilatino λ . Their SUSY transformation read $(M, N, \dots = 0, 1, \dots, 9)$

$$\begin{aligned} \delta_\epsilon \psi_M &= D_M \epsilon + \frac{i}{480} F_{MN_1 \dots N_4} \Gamma^{N_1 \dots N_4} \epsilon + \\ &\quad + \frac{1}{96} G^{MPQ} \Gamma_{MNPQ} \epsilon^* - \frac{3}{32} G_{MNP} \Gamma^{NP} \epsilon^*, \\ \delta_\epsilon \lambda &= i P_M \Gamma^M \epsilon^* - \frac{i}{24} G_{MNP} \Gamma^{MNP} \epsilon^*, \end{aligned} \quad (11.208)$$

where

$$G_{MNP} = f_1(\tau, \bar{\tau})(F_1^{(3)})_{MNP} + f_2(\tau, \bar{\tau})(F_2^{(3)})_{MNP} \quad (11.209)$$

$$P_M = f_3(\tau, \bar{\tau}) \partial_M \tau + f_4(\tau, \bar{\tau}) \partial_M \bar{\tau}, \quad (11.210)$$

for certain coefficient functions $f_a(\tau, \bar{\tau})$ (the $SL(2, \mathbb{R})$ vielbeins) whose effect is to make G_{MNP} and P_M $SL(2, \mathbb{R})$ -invariant and $U(1)_R$ covariant in appropriate representations to make (11.208) fully invariant (see Sect. 4.7.1 of [7] for the explicit expressions⁴⁰). The simplest backgrounds have no fluxes

⁴⁰ The 1-form $P_M dx^M$ is the pull-back to spacetime of the vielbein $(1, 0)$ -form $e d\tau$ of the Poincaré metric on $SL(2, \mathbb{R})/U(1)$ on which $U(1)_R$ acts as $e d\tau \rightarrow e^{i\theta} e d\tau$. The $U(1)_R$ -covariant $SL(2, \mathbb{R})$ -invariant 3-form is $G^{(3)} = \frac{1}{2}(1 + \sigma_3) \mathcal{E}' \Omega F^{(3)}$ with $\mathcal{E} \in SL(2, \mathbb{R})$ a representative of the point $\tau \in SL(2, \mathbb{R})/U(1)$ and $F^{(3)}$ the $SL(2, \mathbb{R})$ doublet of field strengths; cf. Sect. 8.5.

$$G_{MNP} = F_{M_1 \dots M_5} = 0. \quad (11.211)$$

The appropriate treatment of SUSY backgrounds with $\partial_M \tau \neq 0$ requires F-theory, see Sect. 14.3. In this chapter, we focus on the simplest class of backgrounds with

$$G_{MNP} = F_{M_1 \dots M_5} = P_M = 0 \quad (\text{no flux background}). \quad (11.212)$$

The SUSY conditions $\delta_\epsilon \psi_M = \delta_\epsilon \lambda = 0$ then reduce to

$$\partial_\mu \epsilon = D_m \epsilon = 0, \quad \mu = 0, \dots, d-1, \quad m = d, \dots, 9, \quad (11.213)$$

where D_m is the Levi-Civita spin-connection in the internal manifold X_{10-d} . The SUSY of parameter ϵ is unbroken iff the spinor ϵ is *parallel*. **Theorem 11.6** implies that a stationary background with no fluxes which preserves some supersymmetry should be Ricci-flat. The internal manifold X_{10-d} satisfies

$$R_{mn} = 0, \quad (11.214)$$

and the Einstein e.o.m. are automatically satisfied since the energy–momentum tensor T_{mn} vanishes by Eq. (11.212). By **Corollary 11.3** X_{10-d} is (up to finite covers) the product of a flat torus T^k times simply-connected, compact, strict CY, HK, G_2 , or $Spin(7)$ manifolds. If $d \geq 4$ we remain with

$$X_{10-d} = \left(X_m \times T^{10-d-2m} \right) / \Gamma, \quad \Gamma \text{ a finite group of isometries}, \quad (11.215)$$

where X_m is a compact, simply-connected, strict Calabi–Yau of dimension $m \leq 3$. The (complexified) unbroken SUSYs then have the form

$$\epsilon = \epsilon_D \otimes \epsilon_{CY}, \quad (11.216)$$

where ϵ_D is a constant spinor in the flat $D = 10 - 2m$ dimensions and ϵ_{CY} is a parallel spinor in the strict compact m -CY. The number of unbroken supersymmetries is then *twice* the number $N(D)$ (defined in Eq. (8.23)); thus

$$\#(\text{unbroken SUSYs}) = \begin{cases} 32 & \text{for } m = 1 \\ 16 & \text{for } m = 2 \\ 8 & \text{for } m = 3. \end{cases} \quad (11.217)$$

The analysis for Type IIA is similar. Unbroken supersymmetries again correspond to parallel spinors and their number is still given by Eq. (11.217).

Remark 11.3 There are other possibilities for a SUSY-preserving vacuum configuration of the superstring which we shall not explore in this book. Besides the rich subject of *flux* superstring vacua [97, 98], and of SUSY vacua with $d\tau \neq 0$, we have

the possibility of compactifying the 11-dimensional M-theory⁴¹ to four dimensions on a 7-manifold of G_2 holonomy [24, 25] which leads to a 4d effective theory invariant under 4 supercharges [99, 100].

Compactification on a 3-CY

For the rest of this chapter we focus on the most interesting case $m = 3$. The 4d theory we get compactifying the superstring on a given 3-CY X depends on the original 10d theory, Type IIA, or Type IIB. As already mentioned, for most CYs X there exists a mirror geometric 3-CY $\check{X}$ such that

$$X \leftrightarrow \check{X} \text{ has the same effect as } \text{IIA} \leftrightarrow \text{IIB}. \quad (11.218)$$

We shall see in Sect. 12.1 that IIA and IIB become equivalent when compactified on a circle S^1 . Then the compactification of a Type II theory to 3d on the two geometries

$$\mathbb{R}^{2,1} \times X \times S^1 \quad \text{and} \quad \mathbb{R}^{2,1} \times \check{X} \times S^1 \quad (11.219)$$

should produce *equivalent* low-energy theories in 3d.

The 4d/3d Low-Energy Theories The low-energy theory in the four non-compact dimensions should be a 4d supergravity with 8 supercharges, i.e. a 4d $\mathcal{N} = 2$ SUGRA, which after a further compactification on S^1 yields a low-energy 3d $\mathcal{N} = 4$ SUGRA.

As explained in Sect. 8.1, the structure of a SUGRA is determined by its R-symmetry. 4d $\mathcal{N} = 2$ SUGRA⁴² has a gauged R-symmetry

$$U(2)_R \sim SU(2)_R \times U(1)_R, \quad (11.220)$$

while 3d $\mathcal{N} = 4$ SUGRA has R-symmetry

$$\text{Spin}(4)_R \sim SU(2)_h \times SU(2)_t. \quad (11.221)$$

The light supermultiplets of 4d $\mathcal{N} = 2$ supergravity are

- the *gravitational supermultiplet* with the vielbein e_μ^a , 2 Weyl gravitini ψ_μ^i in the fundamental representation of $U(2)_R$, a gauge vector (the “graviphoton”);
- the *vector supermultiplet* with a gauge vector A_μ , 2 Weyl gaugini in the fundamental of $U(2)_R$, and a complex scalar a of charge 2 under $U(1)_R$;
- the *hypermultiplet*: 4 real scalars in the **2** of $SU(2)_R$ and 2 Weyl hyperini of $U(1)_R$ charge +1.

In (11.192), we counted the number of light spacetime d.o.f. in IIA/IIB compactifications on a 3-CY in terms of the BRS-invariant world-sheet vertices. Comparing with the list of $\mathcal{N} = 2$ supermultiplets, we get the massless matter content

⁴¹ M-theory is introduced in Chap. 13.

⁴² See [101–104] or the books [7, 8].

in Type IIA: $h^{1,1}$ vector multiplets, $h^{2,1} + 1$ hypermultiplets;
 in Type IIB: $h^{2,1}$ vector multiplets, $h^{1,1} + 1$ hypermultiplets.

Let us check this world-sheet counting from the viewpoint of the 10d effective SUGRA in the $\mathbb{R}^{3,1} \times X$ geometry. To avoid confusion, we write $F_{10}^{(k)}$ for a 10d k -form field-strength and $F^{(k)}$ for a 4d one.

Type IIB

We consider first the compactification of Type IIB on a strict 3-CY X . X has a *finite* π_1 so that $b_1(X) = 0$, so no 4d vector arises from the 10d metric or 2-forms. 4d vector fields can arise only from the expansion of $F_{10}^{(5)}$ into harmonics of the internal space X . At the linearized level around $F_{10}^{(3)} = H_{10}^{(3)} = 0$ we have⁴³

$$F_{10}^{(5)} \approx (1 + *_{10}) \sum_a \text{Re}(\eta_a) \wedge F_a^{(2)} + (1 + *_{10}) \sum_A \Xi_A \wedge F_A^{(1)}, \quad (11.222)$$

where $\{\eta_a\}$ are harmonic forms yielding a basis of $H^{3,0}(X) \oplus H^{2,1}(X)$ and $\{\Xi_A\}$ are real harmonic forms making a basis of $H^4(X, \mathbb{R})$. By Poincaré duality $\dim_{\mathbb{R}} H^4(X, \mathbb{R}) = h^{1,1}$. Therefore from $F_{10}^{(5)}$ we get $h^{2,1} + 1$ vectors and $h^{1,1}$ scalars. The harmonic expansions of the two field-strength 3-forms, $H_{10}^{(3)}$ and $F_{10}^{(3)}$, give $h^{1,1} + 1$ scalars each: indeed

$$H_{10}^{(3)} = \sum_A \omega_A F_A^{(1)} + H^{(3)}, \quad (11.223)$$

where ω_A is a basis of real harmonic 2-forms, while the last scalar is dual to the 4d NS-NS 2-form $B^{(2)}$

$$d\sigma = *H^{(3)}. \quad (11.224)$$

By Yau theorem, the deformations of the metric, $\delta g_{i\bar{k}}$, δg_{ik} and $\delta g_{\bar{i}\bar{k}}$ which keep it Calabi–Yau yield $h^{1,1} + 2h^{2,1}$ scalars, while $F_{10}^{(1)}$ and Φ give the two scalars which were already present in 10d. In total the number of 4d real scalars in IIB is

$$h^{1,1} + 2(h^{1,1} + 1) + (h^{1,1} + 2h^{2,1}) + 2 \equiv 2h^{2,1} + (4h^{1,1} + 4), \quad (11.225)$$

as predicted by world-sheet BRST invariance and Type IIB GSO projection.

Type IIA

In Type IIA the 4d light bosonic fields are

- from the expansion of $F_{10}^{(4)}$ we get $h^{1,1}$ vectors plus $2h^{2,1} + 2$ scalars. Indeed

$$F_{10}^{(4)} \approx \sum_A \omega_A \wedge F_A^{(2)} + \sum_a \text{Re}(\eta_a) \wedge F_a^{(1)} + \sum_a \text{Im}(\eta_a) \wedge F_a^{(1)'} \quad (11.226)$$

⁴³ The symbol $*_{10}$ stands for the 10d Hodge dual. The Hodge dual in 4d is written simply $*$.

- from $F_{10}^{(2)}$ we get one 4d vector;
- the NS-NS 3-form field-strength $H_{10}^{(3)}$ yields $h^{1,1} + 1$ scalars as in IIB;
- the metric moduli yields $h^{1,1} + 2h^{2,1}$ scalars as in IIB;
- the 10d dilaton Φ yields a last 4d scalar.

In total we have

$$\#(\text{vectors}) = h^{1,1} + 1, \quad \#(\text{scalars}) = 2h^{1,1} + 4(h^{2,1} + 4), \quad (11.227)$$

as expected from the world-sheet SCFT analysis.

The 4d Low-Energy Lagrangian: Preliminaries

We focus on IIB, IIA being related by mirror symmetry. The low-energy theory has $2h^{2,1} + 4h^{1,1} + 4$ light real scalars with kinetic terms of the general form

$$\mathcal{L}_{\text{eff}} = \cdots + \frac{1}{2} \sqrt{-g} G(\phi)_{ab} \partial^\mu \phi^a \partial_\mu \phi^b + \cdots \quad (11.228)$$

for some Riemannian metric $G(\phi)_{ab}$ on the 4d scalars' target space $\mathcal{S}$. When the size of the internal manifold X is large, so its curvatures are small, we may trust the SUGRA approximation: $G(\phi)_{ab}$ may be reliably computed from the classical dimensional reduction of 10d IIB SUGRA. This regime is the weak-coupling limit of the 2d σ -model, whose conformal manifold factorizes into complex and Kähler moduli $\mathcal{C} = \mathcal{M} \times \mathcal{K}$. We now present the corresponding SUGRA result:

Claim 11.3 The *exact* Type IIB light scalars' manifold $\mathcal{S}$ is the product⁴⁴

$$\mathcal{S} = \mathcal{M} \times \mathcal{Q}, \quad (11.229)$$

where the first factor $\mathcal{M}$ is the complex moduli of the Calabi–Yau X equipped with its WP Kähler metric with Kähler form

$$-i \partial \bar{\partial} \log \left(i \int \Omega(s) \wedge \bar{\Omega}(s) \right). \quad (11.230)$$

The factor $\mathcal{Q}$ is an irreducible QK manifold of negative Ricci curvature,⁴⁵ $R_{\alpha\beta} = -\Lambda g_{\alpha\beta}$ with $\Lambda > 0$.

We shall prove **Claim 11.3** below after a lightning review of 4d $\mathcal{N} = 2$ supergravity. Before going to that, we make an important remark. In the 2d SCFT analysis we showed that the geometry of the complex moduli factor $\mathcal{M}$ of the $(2, 2)$ conformal manifold $\mathcal{C}$ was not renormalized by any quantum correction, perturbative or non-perturbative. In that context “quantum correction” meant a correction of the 2d QFT

⁴⁴ Up to finite covers, as always.

⁴⁵ In a certain canonical normalization, Λ is a universal constant depending only on the dimension of $\mathcal{Q}$ [105] or [7].

living on the string world-sheet. These corrections are called α' -*corrections* since the world-sheet action has an overall factor $1/2\pi\alpha'$ which plays the role of $1/\hbar$ in the 2d QFT. In the full string theory there is another, more intrinsic, notion of “quantum corrections” namely the perturbative expansion as a sum over the genera g of the world-sheets (and the non-perturbative completion of this stringy expansion), see Chap. 1. In this expansion the role of the loop-order counting parameter $\hbar$ is played by the string coupling squared $\langle e^{2\Phi} \rangle$. This second class of “quantum corrections” is called *string loop corrections*. We know that the geometry of $\mathcal{M}$ is not affected by the α' -corrections; we shall show that it is also *not renormalized* by any stringy perturbative or non-perturbative correction.

11.7 Lightning Review of 4d $\mathcal{N} = 2$ Supergravity

References for 4d $\mathcal{N} = 2$ SUGRA are [101–104, 106]; for surveys see [7, 8].

In 4d $\mathcal{N} = 2$ supergravity we have two kinds of light scalars: the vector-multiplets’ ones and the hypermultiplets’ ones. At the level of *linear representations* of the SUSY algebra, they are distinguished by their transformations under the R-symmetry group $SU(2)_R \times U(1)_R$. We saw in Sect. 8.1 that in SUGRA the R-symmetry is a local (i.e. gauged) symmetry with a composite operator connection; in particular it is an *exact* symmetry of $\mathcal{L}_{\text{eff}}$. At the *linear* level:

- vector-multiplet scalars are $SU(2)_R$ -invariant and have charge ± 2 under $U(1)_R$;
- hypermultiplet scalars transform in the $\mathbf{2}$ of $SU(2)_R$ and are inert under $U(1)_R$.

The exact statement, valid at the full non-linear level, is obtained by the argument in Note 8.1: the real tangent bundle $T_{\mathbb{R}}\mathcal{S}$ of the scalars’ manifold splits as

$$T_{\mathbb{R}}\mathcal{S} = \mathcal{V}_h \oplus \mathcal{V}_v \rightarrow \mathcal{S} \quad (11.231)$$

where $\mathcal{V}_h$ (resp. $\mathcal{V}_v$) are vector bundles whose fibers transform in a representation of $SU(2)_R \times SO(2)_R$ isotypic⁴⁶ to $(\mathbf{2}, \mathbf{1})$ (resp. $(\mathbf{1}, \mathbf{2})$). The splitting (11.231) is preserved by Levi-Civita parallel transport along $\mathcal{S}$.⁴⁷ The local symmetry $SU(2)_R \times SO(2)_R$ acts on the fibers of $T_{\mathbb{R}}\mathcal{S}$ via the representation

$$R_h \oplus R_v = (\mathbf{2}, \mathbf{1})^{\oplus m} \oplus (\mathbf{1}, \mathbf{2})^{\oplus n} \quad (11.232)$$

with m and n the number of hypermultiplets and vector-multiplets, respectively

$$\text{rank}_{\mathbb{R}}\mathcal{V}_h = 4m, \quad \text{rank}_{\mathbb{R}}\mathcal{V}_v = 2n. \quad (11.233)$$

⁴⁶ W is isotypic to an irreducible representation V , iff it is the direct sum of copies of V : $W = V^{\oplus n}$.

⁴⁷ Since the Lagrangian $\mathcal{L}_{\text{eff}}$ is invariant under the local symmetry R , the holonomy Lie algebra $\mathfrak{hol}$ of the (11.228) is *normalized by the action of G on the scalars*—in plain english: parallel transport in field space is a gauge-equivariant operation.

The holonomy representation ρ normalizes $R_h \oplus R_v$, so it decomposes as

$$\rho = \rho_h \oplus \rho_v, \quad \rho_x: \text{Hol}(\mathcal{S}) \rightarrow \text{End}(\mathcal{V}_x), \text{ for } x = h, v. \quad (11.234)$$

By **Theorem 11.4** the universal cover $\tilde{\mathcal{S}}$ of $\mathcal{S}$ is the product of a manifold $\tilde{\mathcal{Q}}$ of real dimension $4m$ with $T\tilde{\mathcal{Q}} \simeq \mathcal{V}_h$ and one $\tilde{\mathcal{M}}$ of real dimension $2n$ with $T\tilde{\mathcal{M}} \simeq \mathcal{V}_v$

$$\tilde{\mathcal{S}} = \tilde{\mathcal{M}} \times \tilde{\mathcal{Q}}. \quad (11.235)$$

The holonomy groups are then⁴⁸

$$\text{Hol}(\tilde{\mathcal{Q}}) \subseteq N(SU(2)_R)_{SO(4m)} \simeq Sp(1) \cdot Sp(m) \quad (11.236)$$

$$\text{Hol}(\tilde{\mathcal{M}}) \subseteq N(SO(2)_R)_{SO(2n)} \simeq U(n). \quad (11.237)$$

In plain english: $\tilde{\mathcal{M}}$ is Kähler and $\tilde{\mathcal{Q}}$ is quaternionic Kähler. The Ricci curvature of $\tilde{\mathcal{Q}}$ cannot be zero because this will mean that the $Sp(1) \equiv SU(2)_R$ connection is flat which is not possible in SUGRA [7]. Hence the factor $\tilde{\mathcal{Q}}$ is Einstein [23]

$$R_{\alpha\beta} = \Lambda g_{\alpha\beta} \quad (11.238)$$

and irreducible. From SUGRA one shows (Bagger and Witten [105]) that Λ is a universal constant which depends only on the dimension $4m$ of the quaternionic Kähler space. Its precise value depends on the normalization of the metric (see [105]); the crucial point is that Λ is *negative*. This sign may be understood on geometric grounds. The Salamon–Lebrun conjecture [107] states that a quaternionic Kähler manifold with $\Lambda > 0$ is symmetric.⁴⁹ Since the symmetric metrics are rigid, $\Lambda > 0$ would imply a super-strong non-renormalization theorem forbidding all corrections. Computing string amplitudes in backgrounds preserving 8 supersymmetries, one finds several corrections, so we must have $\Lambda < 0$.

The other factor space $\tilde{\mathcal{M}}$ is Kähler. However it cannot be *simply* Kähler—it must carry additional geometric structures—because the kinetic terms of the vector-multiplet scalars are part of the SUSY completion of the vectors’ kinetic terms, and these ones have special geometric structures induced by electromagnetic dualities [7]. These structures are inherited by their superpartners’ space $\tilde{\mathcal{M}}$. Thus the kinetic terms metric on $\tilde{\mathcal{M}}$ should belong to a *special class* of Kähler metrics with a lot of “magical” properties. The “magical” Kähler metrics arising in 4d $\mathcal{N} = 2$ supergravity are dubbed *special Kähler geometries* [7, 9, 12]. The basics of this geometry will be sketched below. Before going to that, let us comment on two important consequences of the factorization of the $\mathcal{N} = 2$ SUGRA scalars’ manifold, Eq. (11.235), for the compactification of the Type IIB superstring on a strict Calabi–Yau manifold X .

⁴⁸ The notation $N(H)_G$ stands for the normalizer of the subgroup $H \subset G$ in the group G .

⁴⁹ A complete Einstein manifold $R_{\alpha\beta} = \Lambda g_{\alpha\beta}$ with $\Lambda > 0$ is *compact* by Myers theorem [108].

A. New Non-Renormalization Theorems In string theory we have two kinds of “quantum corrections”: the stringy loop corrections associated to higher genus world-sheets, which are controlled by the string coupling $e^{2\Phi}$ which vanish asymptotically as $e^\Phi \rightarrow 0$, and the α' corrections, i.e. the quantum correction of the 2d QFT on a fixed world-sheet. The α' corrections are controlled by the v.e.v. of a scalar σ (the “breathing mode”) which plays the role of overall Kähler modulus; these 2d field-theoretic corrections vanish as the volume of the internal Calabi–Yau manifold goes to infinity. We have seen in Sect. 11.6 that the breathing mode σ belongs to the space $\mathcal{K} \subset \mathcal{Q}$, so it is a hypermultiplet scalar in IIB as it is the dilaton Φ , which is universally present. Hence both moduli, Φ, σ , which control the two classes of quantum corrections, belong to the factor space $\mathcal{Q}$. Thus the couplings described by the other factor space $\mathcal{M}$ are Φ - and σ -independent, and cannot receive any Φ or σ dependent quantum correction. Therefore geometry of $\mathcal{M}$ can be computed in the classical limit, where it coincides with the classical geometry of the complex moduli with its canonical WP metric. We conclude that special Kähler geometry in the sense of 4d $\mathcal{N} = 2$ supergravity must be identical to the WP geometry of the moduli space of a family of “abstract” Calabi–Yau 3-folds (possibly “non-geometrical”) as described in “transcendental” Algebraic Geometry [1, 2, 109]. We shall elaborate on this identification between these two, *a priori* unrelated, geometries momentarily.

B. Dimensional Reduction to 3d Consider the compactification from 4d to 3d of our low-energy supergravity $\mathcal{L}_{\text{eff}}$ on a circle S^1 . At low-energy we reduce to a 3d $\mathcal{N} = 4$ SUGRA whose local R-symmetry is $\text{Spin}(\mathcal{N}) \equiv \text{Spin}(4) \simeq SU(2)_h \times SU(2)_t$, 3d $\mathcal{N} = 4$ supergravity has 2 kinds of matter multiplets. First, as in 4d $\mathcal{N} = 2$, we have the hypermultiplet, whose scalars transform (at the linearized level) in the $(2, 1)$ of $\text{Spin}(4)$. The second matter supermultiplet is the *twisted hypermultiplet* whose scalars transform in the $(1, 2)$. de Rham’s theorem implies that the (universal cover of) the scalar manifold is a Riemannian product

$$\tilde{\mathcal{Q}}_h \times \tilde{\mathcal{Q}}_t, \quad (11.239)$$

the two factors being both irreducible quaternionic Kähler manifolds with negative Ricci curvature. In the SUGRA approximation (valid for large radius of the S^1), the QK manifold $\tilde{\mathcal{Q}}_h$ is just the same QK space $\tilde{\mathcal{Q}}$ we had in 4d. The second space $\tilde{\mathcal{Q}}_t$ has real dimension $4(h^{2,1} + 1)$. It contains:

- $2h^{2,1}$ real scalars from the complex moduli space $\mathcal{M}$
- $h^{2,1}$ scalars from the components A_4^a of the matter vectors
- $h^{2,1}$ scalars obtained by dualizing the 3d vector field-strengths $d\sigma^a = *_3 dA^a$
- 2 scalars from the graviphoton: A_4 and $d\sigma = *_3 dA$
- the scalar g_{44} giving the length of S^1
- the scalar dual to the 3d KK vector $g_{\mu 3}$.

From the viewpoint of 3d SUGRA, mirror symmetry interchanges IIA $\leftrightarrow$ IIB, $h^{2,1} \leftrightarrow h^{1,1}$, and acts on the low-energy theory by flipping the two QK factors

$$\tilde{\mathcal{Q}}_h \leftrightarrow \tilde{\mathcal{Q}}_t, \quad (11.240)$$

exchanging hypermultiplets with twisted-hypermultiplets. This was to be expected since IIA and IIB are equivalent when compactified on S^1 ; cf. Sect. 12.1.

Special Kähler Geometry

The extra geometric structure on the vector-multiplet scalars' space $\mathcal{M}$, besides the Kähler one, comes from the physical properties of the vectors' kinetic terms. We review them in an useful language.

Structure of Vectors' Kinetic Terms For a general 4d theory,⁵⁰ *not necessarily supersymmetric*, the vectors' kinetic terms have the form (we leave the Lorentz indices implicit; they are contracted in the obvious way)

$$\mathcal{L}_{\text{eff}} = \cdots - \frac{1}{8\pi} \sqrt{-g} \left(\tau_{ab}(\phi) F_+^a F_+^b + \bar{\tau}_{ab}(\phi) F_-^a F_-^b \right) + \cdots \quad (11.241)$$

where $a, b = 1, \dots, g \equiv \#(\text{vectors})$, and $F_\pm^a$ are (anti-)self-dual projections of the vector field-strengths

$$F_\pm^a = \frac{1}{2} (1 \mp i*) F^a \equiv \frac{1}{2} (1 \mp i*) dA^a, \quad (11.242)$$

and

$$\tau_{ab} = \tau_{ba} \equiv \frac{\theta_{ab}}{2\pi} + \left(\frac{4\pi i}{g^2} \right)_{ab} \quad (11.243)$$

is the complexified gauge-coupling matrix taking value in the Siegel upper-half space

$$\mathcal{H}_g \stackrel{\text{def}}{=} \{ \tau \in \mathbb{C}(g) : \tau = \tau^t, \text{Im } \tau > 0 \} \simeq Sp(2g, \mathbb{R})/U(g). \quad (11.244)$$

$\tau_{ab} \equiv \tau_{ab}(\phi)$ depends on the scalar fields ϕ . By Gauss' law, the magnetic m^a and electric e_a charges are the flux at infinity of the magnetic and electric field-strengths

$$e_a = \int_{S_\infty^2} \frac{G_a}{2\pi} \stackrel{\text{def}}{=} \int_{S_\infty^2} \left(\text{Im } \tau_{ab} * \frac{F^b}{2\pi} + \text{Re } \tau_{ab} \frac{F^b}{2\pi} \right), \quad m^a = \int_{S_\infty^2} \frac{F^a}{2\pi}. \quad (11.245)$$

By Dirac quantization, the charges $(\mathbf{e}, \mathbf{m}) \equiv (e_a, m^b)$ take value in a lattice $\Lambda \simeq \mathbb{Z}^{2g}$ endowed with an integral symplectic pairing:

$$\langle (\mathbf{e}, \mathbf{m}), (\mathbf{e}', \mathbf{m}') \rangle = e_a m'^a - m^a e'_a \in \mathbb{Z}. \quad (11.246)$$

The arithmetic subgroup $Sp(2g, \mathbb{Z}) \subset Sp(2g, \mathbb{R})$ of the $\mathcal{H}_g$ isometry group acts on the electromagnetic fields, fluxes, and charges via the $2g$ -dimensional symplectic representation: its action on the electric/magnetic charges $(\mathbf{e}, \mathbf{m})^t \in \mathbb{Z}^{2g}$ is

⁵⁰ For a more detailed discussion see Chap. 1 of [7]. The original literature is [110, 111].

$$\begin{bmatrix} \mathbf{e}' \\ \mathbf{m}' \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \mathbf{e} \\ \mathbf{m} \end{bmatrix}, \quad \begin{bmatrix} A & B \\ C & D \end{bmatrix} \in Sp(2g, \mathbb{Z}). \quad (11.247)$$

$Sp(2g, \mathbb{Z})$ rotates the electromagnetic duality frame while preserving the Dirac pairing (11.246). The duality frame group $Sp(2g, \mathbb{Z})$ acts on the symmetric matrix $(\boldsymbol{\tau})_{ab} \equiv \tau_{ab}$ by its left action on the symmetric space $Sp(2g, \mathbb{R})/U(h)$

$$\begin{aligned} \boldsymbol{\tau}' &= (A\boldsymbol{\tau} + B)(C\boldsymbol{\tau} + D)^{-1}, \\ \begin{bmatrix} A & B \\ C & D \end{bmatrix} &\in Sp(2g, \mathbb{Z}), \quad \boldsymbol{\tau}, \boldsymbol{\tau}' \in Sp(2g, \mathbb{R})/U(g). \end{aligned} \quad (11.248)$$

The dependence of the vectors' kinetic terms on the scalar fields ϕ^j , $\tau_{ab} = \tau(\phi^j)_{ab}$, then defines a map [13]

$$\tau: \tilde{\mathcal{S}} \rightarrow Sp(2g, \mathbb{R})/U(g), \quad (11.249)$$

where $\tilde{\mathcal{S}}$ is the *universal cover* of the scalars' manifold $\mathcal{S}$.

Note 11.8 The passage to the universal cover $\tilde{\mathcal{S}}$ is required for the matrix τ_{ab} to be univalued. Indeed, when we go along a non-trivial loop in the scalar space $\mathcal{S}$, we may return back to the original point with the electromagnetic frame rotated by a non-trivial element of $Sp(2g, \mathbb{Z})$.

The target space of the gauge-coupling map $\boldsymbol{\tau}$ in Eq. (11.249) has a *deep* geometric interpretation: it is the moduli space of rank- g marked, principally polarized, complex Abelian varieties A [112, 113]. By a *polarization* of A we mean a non-degenerate symplectic pairing

$$- \cdot -: H_1(A, \mathbb{Z}) \times H_1(A, \mathbb{Z}) \rightarrow \mathbb{Z}. \quad (11.250)$$

A polarization is called *principal* if it yields an isomorphism $H^1(A, \mathbb{Z}) \simeq H_1(A, \mathbb{Z})$. A *marking* of A , principally polarized, A is a choice of symplectic $\mathbb{Z}$ -basis in $H_1(A, \mathbb{Z})$

$$\begin{aligned} \alpha_1, \alpha_2, \dots, \alpha_g, \beta_1, \beta_2, \dots, \beta_g &\in H_1(A, \mathbb{Z}) \\ \text{such that } \alpha_a \cdot \alpha_b &= \beta_a \cdot \beta_b = 0, \quad \alpha_a \cdot \beta_b = -\beta_a \cdot \alpha_a = \delta_{ab}. \end{aligned} \quad (11.251)$$

Comparing with Eq. (11.247), we see that Λ gets identified with $H^1(A, \mathbb{Z})$,

$$\begin{aligned} \text{Dirac symplectic lattice of} & \quad \Lambda \longleftrightarrow H^1(A, \mathbb{Z}) \\ \text{electromagnetic charges} & \\ (\mathbf{e}, \mathbf{m})^t & \longleftrightarrow e^a \beta_a - m^a \alpha_a \in H_1(A, \mathbb{Z}) \simeq H^1(A, \mathbb{Z}), \end{aligned} \quad (11.252)$$

the Dirac pairing becomes the polarization, while the different markings of the principally polarized A correspond bijectively to the different electromagnetic duality

frames. Equation (11.249) says that in *any*⁵¹ 4d field theory with g light vectors the (complex) gauge coupling τ_{ab} is a map which associates to a point $p \in \tilde{\mathcal{S}}$, i.e. to a classical vacuum, a marked polarized Abelian variety A_p which is called the *Weil Jacobian* of the gauge coupling in that vacuum.

The cohomology lattice $H^1(A, \mathbb{Z})$ is a topological invariant, hence constant in the moduli space of marked Abelian varieties. A point $p \in Sp(2g, \mathbb{R})/U(g)$ specifies a complex structure on A , hence a Hodge decomposition into $(1, 0)$ and $(0, 1)$ classes

$$H^1(A, \mathbb{Z}) \otimes \mathbb{C} = H^{1,0}(A_p) \oplus H^{0,1}(A_p), \quad H^{0,1}(A_p) = \overline{H^{1,0}(A_p)}, \quad (11.253)$$

which depends on the point p . Through the gauge coupling map $\tau: \phi \mapsto p(\phi) \in \mathcal{H}_g$, each classical vacuum $\phi \in \mathcal{S}$ defines a splitting of the Dirac $\mathbb{C}$ -space $\Lambda_{\mathbb{C}} = \Lambda \otimes \mathbb{C}$

$$\Lambda_{\mathbb{C}} = H^{1,0}(A_{\tau(\phi)}) \oplus H^{0,1}(A_{\tau(\phi)}), \quad H^{0,1}(A_{\tau(\phi)}) = \overline{H^{1,0}(A_{\tau(\phi)})}. \quad (11.254)$$

What is the physical meaning of this Hodge splitting? Let $F = (G_a, F^b)$ be the $2g$ -vector of field-strength 2-forms: its entries are linear combinations of the g curvatures F^a and their dual $*F^a$ which coefficients which depend on the point $p(\phi)$. $H^1(A, \mathbb{Z}) \equiv \Lambda$ is the lattice of the quantized fluxes

$$H^1(A, \mathbb{Z}) \ni \frac{1}{2\pi} \int_{S^2_{\infty}} F = \frac{1}{2\pi} \int_{S^2_{\infty}} F_+ + \frac{1}{2\pi} \int_{S^2_{\infty}} F_-. \quad (11.255)$$

The RHS is the Hodge decomposition. Note that a choice of classical vacuum ϕ defines a linear representation

$$\rho_{\phi}: U(1) \rightarrow Sp(2g, \mathbb{R}) \quad (11.256)$$

such that $e^{i\alpha} \in U(1)$ acts as multiplication by $e^{i\alpha}$ (resp. $e^{-i\alpha}$) on the subspace $H^{1,0}(A_{\tau(\phi)}) \subset \Lambda_{\mathbb{C}}$ (resp. $H^{0,1}(A_{\tau(\phi)}) \subset \Lambda_{\mathbb{C}}$). We leave to the diligent reader to check that this is a *real* symplectic representation of $U(1)$. In other words⁵²:

Fact 11.4 In all Lorentz invariant 4d Lagrangian field theory with g light vectors, the gauge coupling (in a given vacuum) is identified with a real-symplectic representation

$$\rho_{\phi}: U(1) \rightarrow Sp(2g, \mathbb{R}), \quad (11.257)$$

which contains only the $U(1)$ characters $e^{i\alpha}$ and $e^{-i\alpha}$. τ yields a family of such real-symplectic representations parametrized by the cover $\tilde{\mathcal{S}}$ of the scalars' space.

⁵¹ More precisely: in any *Lorentz invariant* 4d field theory.

⁵² **Fact 11.4** is the Deligne viewpoint on 4d gauge couplings [109]. The $U(1)$ is the *Deligne circle* [2]; physically it is the group of unitary automorphisms of the Lorentz algebra $\mathfrak{spin}(3, 1; \mathbb{R})$.

As we shall see below, if the 4d theory happens to be a $\mathcal{N} = 2$ supergravity, the extra structure arising from the Kähler geometry of the vector-multiplets scalars introduces a second, more fundamental, Jacobian, the *Griffiths* one J_p [1].

Geometric Structures in 4d $\mathcal{N} = 2$ SUGRA In 4d $\mathcal{N} = 2$ supergravity coupled to h matter vector-multiplets, we have $g \equiv h + 1$ vectors, the matter ones plus the graviphoton. The R-symmetry implies that the scalars of the vector multiplets live on a Kähler manifold of complex dimension h . The additional geometric structure—besides the Kähler one—on the cover $\widetilde{\mathcal{M}}$ of the vector-multiplet scalar manifold is the gauge-coupling map

$$\tau: \widetilde{\mathcal{M}} \rightarrow \mathcal{H}_{h+1} \stackrel{\text{def}}{=} Sp(2h+2, \mathbb{R})/U(h+1). \quad (11.258)$$

Note 11.9 The vectors’ kinetic terms cannot depend on the scalars in the hypermultiplets, i.e. the gauge couplings τ_{ab} are functions only of the vector-multiplet scalars. Otherwise, by supersymmetry, also the kinetic terms for the vectors’ scalars would depend on the hypermultiplets, and this is forbidden by de Rham’s theorem.

The map (11.258) exists in *all* Lorentz invariant 4d field theories. The new aspect is

Claim 11.4 In $\mathcal{N} = 2$ SUGRA the gauge-coupling map $\tau: \widetilde{\mathcal{M}} \rightarrow \mathcal{H}_{h+1}$ factorizes through the *Griffiths period domain* D_h and the canonical projection⁵³ ϖ

$$\begin{aligned} \widetilde{\mathcal{M}} &\xrightarrow{p} D_h \xrightarrow{\varpi} \mathcal{H}_{h+1} \\ D_h &\stackrel{\text{def}}{=} Sp(2h+2, \mathbb{R})/[U(h) \times U(1)]. \end{aligned} \quad (11.259)$$

Physical Interpretation We explain the factorization (11.259) in physical terms. A point $\phi \in \widetilde{\mathcal{M}}$ specifies a vacuum.⁵⁴ In this vacuum SUSY is linearly realized, and we may use SUSY representation theory. In particular, we have a linear action of the SUSY central charge Z

$$\{Q_\alpha^i, Q_\beta^j\} = \epsilon^{ij} \epsilon_{\alpha\beta} Z \quad (11.260)$$

on the **in/out** asymptotic states. From Eq. (11.260), we see that

$$Z \rightarrow e^{2i\alpha} Z \quad \text{under } e^{i\alpha} \in U(1)_R. \quad (11.261)$$

Since the SUSY algebra is gauged in SUGRA, Z is the charge associated to a gauge symmetry. By Gauss’ law, Z is the flux at infinity of a 2-form field-strength $\mathcal{F}$

⁵³ The subgroup $U(h) \times U(1) \subset Sp(2h+2, \mathbb{R})$ is *compact*, so contained in a maximal compact subgroup $U(h+1) \subset Sp(2h+2, \mathbb{R})$. The canonical projection $\varpi: D_h \rightarrow \mathcal{H}_{h+1}$ maps an equivalence class modulo $U(h) \times U(1)$ to the coarser equivalence class modulo the bigger group $U(h+1)$.

⁵⁴ *A priori* it is just a classical Poincaré invariant vacuum; but now the theory is supersymmetric, and no classical vacuum is lifted by quantum effects. So our vacua are actual quantum vacua.

$$Z = \int_{S^2_\infty} \mathcal{F}, \quad (11.262)$$

where $\mathcal{F}$ is the complex 2-form entering in the SUSY variation of the gravitino [8]

$$\delta\psi_\mu^i = D_\mu\epsilon^i - \frac{1}{4}\epsilon^{ij}\mathcal{F}_{\nu\rho}\gamma^{\nu\rho}\gamma_\mu\epsilon_j + \dots \quad (11.263)$$

(see Chap. 6 of [7] for details). From Eq. (11.261), we see that $\mathcal{F}_{\mu\nu}$ transforms under the chiral symmetry $U(1)_R$ as $\mathcal{F}_{\mu\nu} \rightarrow e^{2i\alpha}\mathcal{F}_{\mu\nu}$, which means that it has a definite chirality, in the sense that it is anti-self-dual

$$\mathcal{F} \equiv \mathcal{F}_-. \quad (11.264)$$

$\mathcal{F}$ is a complex linear combination of the electric and magnetic field strengths $\mathbf{F}$ with coefficients which depend on the particular vacuum ϕ through the coupling $\tau_{ab}(\phi)$. In addition we have the $2h$ complex linear combinations $\mathcal{G}^a \equiv \mathcal{G}_+^a$ of the $\mathbf{F}$, inert under $U(1)_R$, which enter in the SUSY variations of the gauginos. Passing to the corresponding fluxes, we see that, in a given vacuum ϕ , SUSY defines a representation of $U(1)_R$ on the Dirac charge $\mathbb{C}$ -space $\Lambda_{\mathbb{C}} \equiv \Lambda \otimes \mathbb{C}$ containing the characters $e^{2i\alpha}$, $e^{-2i\alpha}$ with multiplicity 1 and the trivial character with multiplicity $2h$. As in Eq. (11.256) this is a real-symplectic representation, i.e.

$$\sigma_\phi: U(1)_R \rightarrow Sp(2h + 2, \mathbb{R}). \quad (11.265)$$

Thus in 4d $\mathcal{N} = 2$ SUGRA a (marked) vacuum $\phi \in \widetilde{\mathcal{M}}$ defines on the Dirac charge $\mathbb{R}$ -space $\Lambda_{\mathbb{R}} \equiv \Lambda \otimes \mathbb{R}$ a real-symplectic representation $\rho_\phi \times \sigma_\phi$ of the group $U(1) \times U(1)_R$ where the first $U(1)$ is the universal one in **Fact 11.4**. Let

$$U(1)_d \subset U(1) \times U(1)_R \quad (11.266)$$

be the diagonal subgroup. In 4d $\mathcal{N} = 2$ **Fact 11.4** has the following extension

Fact 11.5 In a 4d SUGRA coupled to h vector-multiplets, the vector-sector interactions are described by a family of real-symplectic representations

$$\mu_\phi: U(1)_d \rightarrow Sp(2h + 2, \mathbb{R}), \quad e^{i\alpha} \mapsto \rho_\phi(e^{i\alpha})\sigma_\phi(e^{i\alpha}), \quad \phi \in \widetilde{\mathcal{M}}, \quad (11.267)$$

which contains the characters $e^{\pm 3i\alpha}$ with multiplicity 1 and $e^{\mp i\alpha}$ with multiplicity h .

The centralizer of the image of $U(1)_d$ in $Sp(2h + 2, \mathbb{R})$ is $U(1) \times U(h)$; hence the representations (11.267) are parametrized by the domain D_h in Eq. (11.259) [1, 2]. The gauge coupling map τ then factorizes as in (11.259).

Corollary 11.15 In a 4d supergravity coupled to h vector multiplets, the interactions in the gauge sector are fully determined by the period map $p: \widetilde{\mathcal{M}} \rightarrow D_h$.

R-symmetry requires the period map p to satisfy some differential conditions. We sketch the story. By its very definition, the complexified (smooth) tangent bundle to the domain D_h splits in the direct sum of homogeneous vector bundles⁵⁵

$$TD_h \otimes \mathbb{C} = \mathcal{E}_6 \oplus \mathcal{E}_4 \oplus \mathcal{E}_2 \oplus \mathcal{E}_{-2} \oplus \mathcal{E}_{-4} \oplus \mathcal{E}_{-6} \quad (11.268)$$

where the fibers of $\mathcal{E}_q$ transform under the group $U(1)_d$ with charge q . Let $T^{1,0}\widetilde{\mathcal{M}}$ be the holomorphic tangent bundle to $\widetilde{\mathcal{M}}$; by the argument around Eq. (11.231) its fibers transform under $U(1)_R$, hence $U(1)_d$, with charge 2. The couplings in L_{SUGRA} , hence the differential p_* of the map p , should preserve the $U(1)_d$ charge. Hence

Fact 11.6 The map p satisfies the differential relation

$$p_*(T^{1,0}\widetilde{\mathcal{M}}) \subseteq \mathcal{E}_2. \quad (11.269)$$

This is the full story. Any map p in Eq. (11.259) which satisfies the differential relation (11.269) defines the vector-multiplet couplings of a formally consistent $\mathcal{N} = 2$ supergravity, that is, a special Kähler geometry. To fully specify the supergravity one needs to give, in addition, the hypermultiplet couplings and the YM gaugings. Equation (11.269) says, in particular, that p is a *holomorphic map* with respect to the standard complex structure of D_h where the $(1, 0)$ vectors are the sections of the sub-bundle $\bigoplus_{q>0} \mathcal{E}_q \subset TD_h \otimes \mathbb{C}$. The map p determines the special Kähler metric on $\widetilde{\mathcal{M}}$ as follows. $\mathcal{E}_6$ is a homogeneous holomorphic line bundle over D_h with a canonical metric so has a canonical Chern connection. The Kähler form ω of the special Kähler metric is the $(1, 1)$ form

$$\omega = \frac{1}{2} p^* \Theta \quad (11.270)$$

where Θ is the curvature of the canonical connection on the line bundle $\mathcal{E}_6 \rightarrow D_h$. Since Θ is a universal $(1, 1)$ -form on D_h (see Eq. (11.324) for its explicit expression), the special Kähler metric is captured by the map p which also gives the vectors' kinetic terms. Further details about 4d $\mathcal{N} = 2$ SUGRA are discussed below in the context of compactifications of Type IIB on Calabi–Yau 3-folds.

11.8 The Low-Energy Theory of Type IIB on a 3-CY X

We assume that our 4d $\mathcal{N} = 2$ SUGRA gives the low-energy description of Type IIB compactified on a (family of) Calabi–Yau 3-folds X .⁵⁶ Our method to get the

⁵⁵ For homogeneous bundles over Griffiths period domains see [7] or the original literature [114].

⁵⁶ As a matter of notation, X represents the underlying smooth space *stripped* of the complex structure. X_s for $s \in \mathcal{M}$ stands for the manifold endowed with the complex structure s .

*exact*⁵⁷ low-energy Lagrangian $\mathcal{L}_{\text{eff}}$ is to identify the geometric structures of $\mathcal{N} = 2$ SUGRA, singled out in Sect. 11.7, with the geometric structures implied by the Algebraic Geometry of CY 3-folds [11, 12, 115, 116]. Our conclusions will be valid for all $\mathcal{N} = 2$ SUGRAs, whether or not they arise from a Calabi-Yau.

We know from the local Torelli theorem that the local geometry at a point $\phi \in \mathcal{M}$ is described by the Hodge structure of the Calabi-Yau X_ϕ , i.e. by the decomposition of the fixed vector space

$$H^3(X, \mathbb{Z}) \otimes_{\mathbb{Z}} \mathbb{C} \equiv H^3(X, \mathbb{C}) = \bigoplus_{p+q=3} H^{p,q}(X_\phi), \quad H^{q,p}(X_\phi) = \overline{H^{p,q}(X_\phi)}. \quad (11.271)$$

The lattice $H^3(X, \mathbb{Z})$ has a *polarization*, i.e. a skew-symmetric integral pairing

$$Q(\alpha, \beta) = -Q(\beta, \alpha) \in \mathbb{Z}, \quad \text{for } \alpha, \beta \in H^2(X, \mathbb{Z}), \quad (11.272)$$

given by minus the intersection form

$$Q(\alpha, \beta) = - \int_X \alpha \wedge \beta \in \mathbb{Z}, \quad (11.273)$$

which extends by linearity to a non-degenerate skew-symmetric form on $H^3(X, \mathbb{C})$. If $V \subset H^3(X, \mathbb{C})$ is a subspace, we set

$$V^\perp \stackrel{\text{def}}{=} \left\{ w \in H^3(X, \mathbb{C}) : Q(v, w) = 0 \ \forall v \in V \right\}. \quad (11.274)$$

The *first Riemann bilinear relation* reads [1, 2, 20, 28]

$$\begin{aligned} Q(\alpha, \beta) &= 0 \text{ for } \alpha \in H^{p,q}(X_\phi), \beta \in H^{p',q'}(X_\phi) \\ &\text{unless } p + p' = q + q' = 3. \end{aligned} \quad (11.275)$$

Consider the Hermitian form (anti-linear in the second argument)

$$\langle \alpha, \beta \rangle \stackrel{\text{def}}{=} Q(C_\phi \alpha, \bar{\beta}), \quad \alpha, \beta \in H^3(X, \mathbb{C}), \quad (11.276)$$

where $C_\phi : H^3(X, \mathbb{C}) \rightarrow H^3(X, \mathbb{C})$ is the (real) *Weil operator*

$$C|_{H^{p,q}(X_\phi)} = i^{p-q}. \quad (11.277)$$

The *second Riemann bilinear relation* [1, 2, 20, 28] states that (11.276) is positive-definite. Indeed, for harmonic forms α, β

⁵⁷ The extend $\mathcal{L}_{\text{eff}}$ is exact is specified by the SUSY non-renormalization theorems.

$$\langle \alpha, \beta \rangle = \int_X \alpha \wedge * \bar{\beta}. \quad (11.278)$$

We stress that the dependence of $\langle -, - \rangle$ on the moduli is only through C_ϕ , that is, through the Hodge decomposition (11.271).

By a *marking* of X we mean a choice of canonical symplectic basis (A_I, B^J) , $(I, J = 0, \dots, h^{2,1})$ of $H_3(X, \mathbb{Z})$ (choice of A and B 3-cycles) with intersection form

$$A_I \cdot A_J = B^I \cdot B^J = 0, \quad A_I \cdot B^J = -B^J \cdot A_I = \delta_I^J. \quad (11.279)$$

Any two markings differ by the action of an element of $Sp(2h^{2,1} + 2, \mathbb{Z})$

$$\begin{bmatrix} A'_I \\ B'^J \end{bmatrix} = \begin{bmatrix} a_I^K & b_{IL} \\ c^{JK} & d^J_L \end{bmatrix} \begin{bmatrix} A_K \\ B^L \end{bmatrix} \quad \begin{bmatrix} a_I^K & b_{IL} \\ c^{JK} & d^J_L \end{bmatrix} \in Sp(2h^{2,1} + 2, \mathbb{Z}). \quad (11.280)$$

Comparing with the decomposition (11.222) of the 10d 5-form flux $F_{10}^{(5)}$ in internal harmonic 3-forms and 4d electric and magnetic fields-strengths, we see that a change in the marking of X has the same effect as a rotation of the electromagnetic duality frame, which is a symplectic rotation of the electric/magnetic field strengths preserving the Dirac quantization of fluxes

$$\begin{bmatrix} G'_{\mu\nu I} \\ F'^J_{\mu\nu} \end{bmatrix} = \begin{bmatrix} a_I^K & b_{IL} \\ c^{JK} & d^J_L \end{bmatrix} \begin{bmatrix} G_{\mu\nu K} \\ F^L_{\mu\nu} \end{bmatrix} \quad (11.281)$$

where the 2-forms $G_{\mu\nu I}$, $F^J_{\mu\nu}$ are related to the quantized electric and magnetic charges, e_I, m^J as in Eq. (11.245). The space which parametrizes all possible Hodge decompositions of the form (11.271) with

$$\dim_{\mathbb{C}} H^{3,0} = \dim_{\mathbb{C}} H^{0,3} \equiv h^{3,0} = 1, \quad \dim_{\mathbb{C}} H^{2,1} = \dim_{\mathbb{C}} H^{1,2} \equiv h^{2,1}, \quad (11.282)$$

and consistent with the 2 Riemann relations is the reductive homogeneous space

$$Sp(2h^{2,1} + 2, \mathbb{R})/[U(1) \times U(h^{2,1})] \equiv D_{h^{2,1}}, \quad (11.283)$$

by the argument before **Corollary 11.15** applied to the $U(1)$ action on $H^3(X, \mathbb{C})$

$$\rho_\phi(e^{i\alpha})|_{H^{p,q}(X_\phi)} = e^{i(q-p)\alpha}. \quad (11.284)$$

We conclude that the Hodge decomposition yields a map from the complex moduli space $\widetilde{\mathcal{M}}$ of *marked* Calabi–Yau to the Griffiths period domain $D_{h^{2,1}}$

$$p: \widetilde{\mathcal{M}} \rightarrow Sp(2h^{2,1} + 2, \mathbb{R})/[U(1) \times U(h^{2,1})] \quad (11.285)$$

called the *Griffiths period map* [1, 2]. It should be compared with the map p in Eq. (11.259) present in all 4d $\mathcal{N} = 2$ SUGRA on physical grounds. The Lagrangian $\mathcal{L}_{\text{eff}}$ of the effective 4d $\mathcal{N} = 2$ supergravity is determined by the period map p . In particular the gauge-coupling τ is obtained by composing the Griffiths period map of X with the canonical projection ϖ as in Eq. (11.259), while the scalars' Kähler metric is determined as in Eq. (11.270).

In Algebraic Geometry, the period map p satisfies a set of differential relations called the *(Griffiths) infinitesimal period relations* (IPR) [1, 2].

Claim 11.5 The IPR are *exactly* the SUSY relations (11.269).

Proof From KS theory, a form representing a class in $H^{p,q}(X_s)$ has the form (cf. Lemma 11.1)

$$\Psi(s)^{(p,q)} \equiv \Psi(s)_{i_1 \dots i_p \bar{k}_1 \dots \bar{k}_q} \omega(s)^{i_1} \wedge \omega(s)^{i_p} \wedge \overline{\omega(s)}^{\bar{k}_1} \wedge \dots \wedge \overline{\omega(s)}^{\bar{k}_q}, \quad (11.286)$$

where $\Psi(s)_{i_1 \dots i_p \bar{k}_1 \dots \bar{k}_q}$ are suitable coefficients and

$$\omega(s)^j = dz^j - \phi(s)_{\bar{k}}^j d\bar{z}^{\bar{k}}, \quad (11.287)$$

with $\phi(s)_{\bar{k}}^j$ the KS vector which depends holomorphically on the coordinates s^a of the deformation space $\mathcal{S}$. Taking derivatives of (11.286) we see that

$$\partial_{s_a} \Psi(s)^{(p,q)} \in H^{p,q}(X_s) \oplus H^{p-1,q+1}(X_s) \quad (11.288)$$

$$\partial_{\bar{s}_a} \Psi(s)^{(p,q)} \in H^{p,q}(X_s) \oplus H^{p+1,q-1}(X_s), \quad (11.289)$$

which are equivalent to (11.269) under the identification of the $U(1)_d$ charge of a vector in $\Lambda_{\mathbb{C}}$ with $q - p$ for a vector in $H^3(X, \mathbb{C})$ ($\equiv \Lambda_{\mathbb{C}}$ by Eq. (11.222)). To see the equivalence, note that the complexified tangent space to D_h consists of infinitesimal variations of the decomposition of linear $\mathbb{C}$ -space $\Lambda_{\mathbb{C}} = \bigoplus_{p,q} H^{p,q}$, and hence is contained in $\bigoplus_{k \neq 0} \bigoplus_{p+q=3} \text{Hom}(H^{p,q}, H^{p-k,q+k})$, while its direct summands $\mathcal{E}_{2k}$ in (11.268) are defined as

$$\mathcal{E}_{2k} = T_{\mathbb{C}} D_h \bigcap \bigoplus_{p+q=3} \text{Hom}(H^{p,q}, H^{p-k,q+k}). \quad (11.290)$$

□

The Explicit Form of $\mathcal{L}_{\text{eff}}$

To write the vector-multiplet sector effective Lagrangian explicitly, we observe that to fix the Hodge decomposition (11.271) it is enough to specify the 2 subspaces

$$F^3 \equiv H^{3,0} \subset H^3(X, \mathbb{C}) \quad \text{and} \quad F^2 = H^{3,0} \oplus H^{2,1} \subset H^3(X, \mathbb{C}) \quad (11.291)$$

since setting

$$F^1 = (F^3)^{\perp} \quad \text{and} \quad F^0 \equiv H^3(X, \mathbb{C}) \quad (11.292)$$

we recover all Hodge subspaces $H^{p,q}$ as

$$H^{p,q} = F^p \cap \overline{F}^q. \quad (11.293)$$

We return to our family of CYs in Sect. 11.1.4. We have a holomorphic fibration

$$\mathcal{X} \rightarrow \widetilde{\mathcal{M}} \quad (11.294)$$

whose fibers are marked polarized 3-CY. We consider a holomorphic 3-form $\Omega(s)$ on the total space $\mathcal{X}$ which restricts to a non-zero holomorphic $(3, 0)$ -form on each fiber. By **Lemma 11.1**

$$\Omega(s) = \Omega(s)_{ijk} \omega(s)^i \wedge \omega(s)^j \wedge \omega(s)^k \quad (11.295)$$

with $\omega(s)^i$ as in (11.287). Taking derivatives as in (11.288) we get⁵⁸

$$\Omega(s)| \in H^{3,0} \equiv F^3 \quad (11.296)$$

$$\partial_{s^a} \Omega(s)| \in H^{3,0} \oplus H^{2,1} \equiv F^2 \quad (11.297)$$

$$\partial_{s^a} \partial_{s^b} \Omega(s)| \in H^{3,0} \oplus H^{2,1} \oplus H^{1,2} \equiv F^1 \quad (11.298)$$

$$\partial_{s^a} \partial_{s^b} \partial_{s^c} \Omega(s)| \in H^{3,0} \oplus H^{2,1} \oplus H^{1,2} \oplus H^{0,3} \equiv F^0, \quad (11.299)$$

while, as a corollary to the local Torelli theorem,

$$F^3 = \text{span}(\Omega(s)|), \quad F^2 = \text{span}(\Omega(s)|, \partial_{s^a} \Omega(s)|) \quad (11.300)$$

where the spaces F^1, F^0 are as in (11.292). Using Eq. (11.293), we reconstruct the full Hodge decomposition at each point $s \in \widetilde{\mathcal{M}}$ from $\Omega(s)$ and its first derivatives.

Corollary 11.16 *For a Calabi–Yau 3-fold, to specify the full Griffiths period map*

$$p: \widetilde{\mathcal{M}} \rightarrow D_{h^{2,1}} \quad (11.301)$$

it is enough to give a nowhere zero holomorphic section

$$\Omega(s) \in \Gamma(\widetilde{\mathcal{M}}, \mathcal{F}^3) \quad (11.302)$$

where $\mathcal{F}^3 \rightarrow \widetilde{\mathcal{M}} \rightarrow D_{h^{2,1}}$ is the holomorphic line bundle with fiber

$$\mathcal{F}_s^3 \equiv H^{3,0}(X_s). \quad (11.303)$$

$\Omega(s)$ is unique only up to a holomorphic rescaling

$$\Omega(s) \rightarrow e^{f(s)} \Omega(s) \quad (11.304)$$

with $f(s)$ holomorphic. The transformation (11.304) is a kind of gauge symmetry (physicists like to call it “Kähler gauge symmetry”). Physical observables cannot depend on the choice of $\Omega(s)$. To write explicitly $\Omega(s)$, we use its *A- and B-periods* (with respect to any chosen marking of X)

⁵⁸ A vertical bar $|$ stands for restriction of the differential form to the CY fiber.

$$X^I(s) = \int_{A_I} \Omega(s), \quad F_J(s) = \int_{B^J} \Omega(s). \quad (11.305)$$

We see $(X^I(s), F_J(s))$ as a vector in $\mathbb{C}^{2(h^{2,1}+1)}$ which is never zero (since $\Omega(s)$ is non-zero). Under a “gauge” transformation $\Omega(s) \rightarrow e^{f(s)}\Omega(s)$

$$(X^I(s), F_J(s)) \rightarrow e^{f(s)}(X^I(s), F_J(s)) \quad (11.306)$$

so, for each $s \in \widetilde{\mathcal{M}}$, the point

$$(X^0(s) : X^1(s) : \cdots : X^{h^{2,1}} : F_0(s) : F_1(s) : \cdots : F_{h^{2,1}}) \in \mathbb{P}^{2h^{2,1}+1} \quad (11.307)$$

is well-defined independently of the choice of $\Omega(s)$ (but depends on the marking).

Corollary 11.17 *We have a well-defined map*

$$\zeta : \widetilde{\mathcal{M}} \rightarrow \mathbb{P}^{2h^{2,1}+1} \quad (11.308)$$

whose first prolongation is the Griffiths period map p (cf. **Corollary 11.16**).

The map ζ specifies which complex line in the complex Dirac space $\Lambda_{\mathbb{C}} \cong \mathbb{C}^{2h^{2,1}+2}$ corresponds to the SUSY central charge Z in each vacuum $\phi \in \widetilde{\mathcal{M}}$, i.e. the one-dimensional subspace of $\Lambda_{\mathbb{C}}$ on which the R-symmetry $U(1)_R$ acts through the character $e^{i\alpha} \mapsto e^{2i\alpha}$, see discussion after Eq. (11.260).

The Griffiths period map p satisfies the IPR (11.269). These relations can be rewritten as differential conditions on the more basic map ζ . To write the differential relations in a convenient way, the crucial observation [115] is that an odd dimensional complex projective space $\mathbb{P}^{2m+1}$ carries a canonical *holomorphic contact structure*.⁵⁹ Let $(z^0 : z^1 : \cdots : z^{2m+1})$ be homogeneous coordinates in $\mathbb{P}^{2m+1}$ and set

$$Q(z, z') = \Omega_{ab} z^a z'^{b'} \quad (11.309)$$

where $\Omega \equiv i\sigma_2 \otimes \mathbf{1}_{m+1}$ is the usual $(2m+2) \times (2m+2)$ symplectic matrix. The holomorphic 1-form

$$\kappa = Q(z, dz) \equiv \sum_{k=0}^{m+1} (z^k dz^{k+m+1} - z^{k+m+1} dz^k) \quad (11.310)$$

defines a canonical holomorphic contact structure on $\mathbb{P}^{2m+1}$ [19]: indeed the holomorphic line sub-bundle of $T^*\mathbb{P}^{2m+1}$ generated by κ is invariant under an overall rescaling of the homogeneous coordinates $z^a \rightarrow \lambda z^a$

$$Q(\lambda z, d(\lambda z)) = \lambda^2 Q(z, dz) + \lambda Q(z, z) d\lambda = \lambda^2 Q(z, dz) \equiv \lambda^2 \kappa \quad (11.311)$$

⁵⁹ For the notion of contact structure see [19] or [117, 118].

(since $Q(z, z) \equiv 0$), while (setting $z^0 = 1$)

$$\kappa \wedge (d\kappa)^m = 2^m m! dz^{m+1} \wedge \left(\bigwedge_{k=1}^m dz^k \wedge dz^{m+k+1} \right) \neq 0. \quad (11.312)$$

Definition 11.4 A holomorphic Legendre submanifold $\iota: K \rightarrow M$ of a holomorphic contact manifold M of complex dimension $2m + 1$ and contact holomorphic form κ is a holomorphic submanifold $L \subset M$ of complex dimension m such that

$$\iota^* \kappa = 0. \quad (11.313)$$

We know from the Hamilton–Jacobi (HJ) formulation of Classical Mechanics [119, 120] that the *generic*⁶⁰ Legendre submanifold $L \subset \mathbb{P}^{2m+1}$, written in homogeneous coordinates, is the graph of the gradient of a holomorphic function, the “HJ action” $F(z^I)$, of the first $m + 1$ homogeneous coordinates z^I ($I = 0, 1, \dots, m$)

$$(z^0 : z^1 : \dots : z^m : \partial_{z^0} F : \partial_{z^1} F : \dots : \partial_{z^m} F) \subset \mathbb{P}^{2m+1}, \quad (11.314)$$

with $F(z^I)$ homogeneous of degree 2, so that the point in $\mathbb{P}^{2m+1}$ is invariant under the overall rescaling $z^I \rightarrow \lambda z^I$. The image in $\mathbb{P}^{2m+1}$ depends only on the m ratios

$$\frac{z^1}{z^0}, \frac{z^2}{z^0}, \dots, \frac{z^m}{z^0} \quad (11.315)$$

which are local coordinates in the submanifold L . We sketch a proof of (11.314).

Proof (We set $F_I \equiv \partial_{z^I} F$; repeated indices are summed over). We have

$$\iota^* \kappa = \iota^* Q(z, dz) = z^I dF_I - F_I dz^I = d(z^I F_I) - 2F_I dz^I \equiv d(z^I F_I - 2F) = 0, \quad (11.316)$$

since $z^I F_I \equiv 2F$ because F is homogeneous of degree 2. $\square$

Theorem 11.11 (Bryant–Griffiths [115]) **(1)** For a family of (marked) Calabi–Yau 3-folds $\{X_s\}_{s \in \widetilde{\mathcal{M}}}$ the map

$$\zeta: \widetilde{\mathcal{M}} \rightarrow \mathbb{P}^{2h^{2,1}+1} \quad (11.317)$$

is a holomorphic Legendre submanifold.

(2) More generally, for all holomorphic Legendre submanifold

$$\zeta: L \hookrightarrow \mathbb{P}^{2h^{2,1}+1}, \quad (11.318)$$

its first prolongation

$$p: L \rightarrow D_{h^{2,1}} \quad (11.319)$$

satisfies the infinitesimal period relations.

⁶⁰ That is, satisfying the obvious transversality conditions.

Proof We have

$$\zeta^* \kappa = Q(\Omega(s), d\Omega(s)) = 0 \quad (11.320)$$

by the first Riemann relation. $\square$

Corollary 11.18 *For a generic marking of the Calabi–Yau 3-fold, the B -periods (11.305) of the $(3, 0)$ form Ω , as functions of the A -periods X^I , are given by*

$$\int_{B_J} \Omega \equiv F_J = \partial_{X^J} F(X^I) \quad (11.321)$$

for a holomorphic function $F(X^I)$, homogeneous of degree 2, called the pre-potential.

Summary of 4d $\mathcal{N}=2$ SUGRA: *Locally in field-space* the structure of the vector-multiplet couplings in 4d $\mathcal{N} = 2$ SUGRA is as follows:⁶¹

- (a) Any holomorphic function $F(X^I)$, homogenous of degree 2 in the variables $X^0, X^1, \dots, X^h$, defines (locally) a holomorphic Legendre submanifold

$$L \equiv (X^I, \partial_I F) \subset \mathbb{P}^{2h+1} \quad (11.322)$$

whose first prolongation defines a holomorphic map, satisfying the IPR,

$$p: L \rightarrow D_h, \quad (11.323)$$

which fixes all couplings in the vector sector of an (ungauged) 4d $\mathcal{N} = 2$ SUGRA coupled to h vector-multiplets. $F(X^I)$ is called the *pre-potential*.

- (b) Conversely, all 4d $\mathcal{N} = 2$ SUGRAs have locally in field space this form for some pre-potential $F(X^I)$.
- (c) The ratios $z^1 = X^1/X^0, z^2 = X^2/X^0, \dots, z^h = X^h/X^0$ are holomorphic *local* coordinates on the Legendre manifold L called the *special coordinates*.
- (d) The gauge coupling is $\tau = \varpi \circ p$ with $\varpi: D_h \rightarrow \mathcal{H}_{h+1}$ the projection.
- (e) The Legendre complex h -fold L is a holomorphic cover of the Kähler space $\mathcal{M}$ where the vector-multiplet scalars take value. In other words $\mathcal{M} = L/\Gamma$ for some discrete group Γ of isometries.
- (f) The Kähler form of $\mathcal{M}$ (pulled back to $\mathcal{L}$) is the curvature of the $p^*\mathcal{F}^3 \rightarrow \mathcal{M}$ holomorphic bundle, whose explicit expression in terms of the derivatives of the pre-potential $F(X^I)$ is

$$\omega = -i\partial\bar{\partial} \log \left(i \int_X \Omega(s) \wedge \bar{\Omega}(s) \right) = -i\partial\bar{\partial} \log (iX^I \bar{F}_I - iF_I \bar{X}^I). \quad (11.324)$$

When the SUGRA arises from an actual family of *geometric* CYs, (11.324) is the Weil–Petersson Kähler form of its complex moduli space.

⁶¹ The following statements hold for some *sufficiently generic* duality frame. The choice of a “good” frame may be obstructed if the SUGRA is gauged.

The Kähler potential K has the local expression (cf. (11.324))

$$e^{-K(s, \bar{s})} |X^0|^2 = \|\Omega(s)\|^2 \equiv (iX^I \bar{F}_I - iF_I \bar{X}^I) \quad (11.325)$$

e^{-K} being the fiber metric on the line bundle $\mathcal{F}^3$. The web of geometric structures, relations, and compatibility conditions i(a)–(f) is called *special (Kähler) geometry*.

When our 4d $\mathcal{N} = 2$ SUGRA is the low-energy limit of IIB on a CY 3-fold X , the period map p which determines its vector-multiplet couplings is the Algebro-Geometric period map which describes the VHS on its complex moduli space [1–3]. The compactification of IIA on the CY manifold X leads to the same story as for IIB with X replaced by its mirror Calabi–Yau $\check{X}$.

“Yukawa Couplings”

In special geometry, there is an important tensorial invariant which in the old math literature was called *the cubic form of the infinitesimal variations of Hodge structure* [1, 115]. Nowadays mathematicians call it the “Yukawa couplings” [2].

We write $T\mathcal{M}$ for the holomorphic tangent space to the complex moduli space⁶² $\mathcal{M}$ and $\mathcal{F}^p \rightarrow \mathcal{M}$ for the holomorphic vector bundle whose fiber at s is

$$\mathcal{F}_s^p \equiv \bigoplus_{p' \geq p} H^{p', 3-p'}(X_s). \quad (11.326)$$

Note that $\mathcal{F}^3$ is a line bundle.

Proposition 11.5 *On the complex moduli $\mathcal{M}$ of a Calabi–Yau 3-fold, we have a symmetric cubic form $\mathcal{Y}$ on $T\mathcal{M}$ with values in the line bundle $(\mathcal{F}^3)^{-2}$*

$$\mathcal{Y}: \mathcal{F}^3 \otimes \odot^3 T\mathcal{M} \rightarrow \mathcal{F}^0/\mathcal{F}^1 \simeq (\mathcal{F}^3)^\vee, \quad \mathcal{Y} \in \Gamma(\odot^3 T^*\mathcal{M} \otimes (\mathcal{F}^3)^2) \quad (11.327)$$

called “Yukawa coupling”, given by $\partial_{s^i} \partial_{s^j} \partial_{s^k} \Omega(s) \bmod \mathcal{F}^1$, i.e.

$$\mathcal{Y} = Q(\partial_{s^i} \partial_{s^j} \partial_{s^k} \Omega(s), \Omega(s)) ds^i \odot ds^j \odot ds^k \in \Gamma(\odot^3 T^*\mathcal{M} \otimes \mathcal{F}^2). \quad (11.328)$$

Proof Equations (11.296)–(11.299) and the first Riemann bilinear relation. $\square$

The “Yukawa couplings” have a simple expression in terms of the *special coordinates* z^i ($i = 1, \dots, h$). Let $F(X^I) = (X^0)^2 f(X^i/X_0) \equiv (X^0)^2 f(z^i)$ be the superpotential: the periods of $\Omega(s)$ are (here $f_i \equiv \partial_{z^i} f$, $f_{ij} \equiv \partial_{z^i} \partial_{z^j} f$, etc.)

$$\Xi = X^0(1, z^1, \dots, z^h, 2f - z^i f_i, f_1, \dots, f_h) \quad (11.329)$$

$$\Xi_{jkl} = X^0(0, 0, \dots, 0, -f_{jkl} - z^i f_{ijkl}, f_{1jkl}, \dots, f_{hjkl}), \quad (11.330)$$

⁶² Since the considerations are strictly local, we do not distinguish between $\mathcal{M}$ and its cover $\widetilde{\mathcal{M}}$.

so that

$$\mathcal{Y}_{jkl} \equiv \frac{1}{(X^0)^2} \Xi_{jkl}^t \Omega \Xi = f_{ijk} \equiv X^0 \partial_{X^i} \partial_{X^j} \partial_{X^k} F(X^I). \quad (11.331)$$

The last expression for $\mathcal{Y}_{ijk}$ as the 3rd derivative of the pre-potential in special coordinates makes sense for *all* special Kähler geometries regardless if they originate from an actual family of geometric CYs. Thus the “Yukawa couplings” are natural tensor invariants of all special geometries.

Exercise 11.5 Show that the Riemann tensor in special geometry has the form

$$R_{i\bar{j}k\bar{l}} = -G_{i\bar{j}} G_{k\bar{l}} - G_{i\bar{l}} G_{k\bar{j}} + e^{2K} \mathcal{Y}_{ikm} \bar{\mathcal{Y}}_{\bar{j}\bar{l}\bar{m}} G^{\bar{m}m}. \quad (11.332)$$

Additional details on special geometry may be found in [9, 12]. For its relation with tt^* geometry see [4, 13, 93].

Why the Name “Yukawa Coupling”? When we use the CY space X to compactify the heterotic $E_8 \times E_8$ string we get a low-energy 4d $\mathcal{N} = 1$ SUGRA coupled to a E_6 gauge theory with a net number $(h^{2,1} - h^{1,1}) = -\chi(X)/2$ of chiral generations in the **27** of E_6 ; cf. Sect. 11.5. We see in (11.203) that the 4d chiral superfield scalars $\Phi^{a,r}$ in the **27** of E_6 arise from the harmonic expansion of the 10d gauge fields as

$$A_{[ij],\bar{k}}^a = \epsilon_{ijl} \Phi^{a,r} (\phi(s)_r)_k^l + \dots \quad (11.333)$$

with $\{\phi(s)_r\}$ a basis of the space $H^1(X_s, TX_s)$ of infinitesimal complex deformations. In view of Eq. (11.270) (and local Torelli) we can choose the basis so that

$$\partial_{s^r} \Omega(s) = \phi(s)_r \rfloor \Omega(s) \bmod H^{3,0}(X_s). \quad (11.334)$$

Their fermionic partners $\chi_\alpha^{a,r}$ arise in a similar way: the 10d gluino λ^a is in the **16** of $SO(9, 1)$ which decomposes as $(\mathbf{2}_+, \mathbf{4}) \oplus (\mathbf{2}_-, \bar{\mathbf{4}})$ of $SO(3, 1) \times SO(6)$; under the subgroup $SU(3) \subset SO(6)$ $\bar{\mathbf{4}} \rightarrow \mathbf{1} \oplus \bar{\mathbf{3}}$. The component in **3** is a 4d Weyl fermion times an internal $(0, 1)$ form and

$$\lambda_{[ij],\bar{k}\alpha}^a = \epsilon_{ijl} \lambda_\alpha^{a,r} (\phi(s)_r)_k^l + \dots \quad (11.335)$$

($\alpha = 1, 2$ is a 4d Weyl spinor index). The 4d effective superpotential arises from a cubic term in the internal components of the gauge fields

$$\begin{aligned} \int_X \text{tr}(\bar{\lambda} \Gamma^{\bar{k}} [A_{\bar{k}} \lambda]) &\rightsquigarrow t_{abc} \Phi^{a,j} \lambda^{b,k} \lambda^{c,l} \int_X \phi(s)_j \rfloor \phi(s)_k \rfloor \phi(s)_l \rfloor \Omega(s) \wedge \Omega(s) = \\ &= t_{abc} \Phi^{a,j} \lambda^{b,k} \lambda^{c,l} \int_X \partial_{s^j} \partial_{s^k} \partial_{s^l} \Omega(s) \wedge \Omega(s) = t_{abc} \Phi^{a,j} \lambda^{b,k} \lambda^{c,l} \mathcal{Y}_{jkl}, \end{aligned} \quad (11.336)$$

where t_{abc} is the unique symmetric invariant three-tensor in the fundamental of E_6 . $\mathcal{Y}_{jkl}$ are now Yukawa couplings in the standard physical sense for the 27 generations. The fact that the Yukawa couplings are sections of a holomorphic bundle rather than complex numbers is related to the fact that the superpotential $\mathcal{W}(X^a)$ in $\mathcal{N} = 1$ supergravity is a section of a holomorphic line bundle, not a holomorphic function (as in *rigid* $\mathcal{N} = 1$ SUSY), on the manifold of chiral superfields [121].

11.9 The Hypermultiplet Sector. c -Map

The hypermultiplet sector of the low-energy effective theory of Type IIB compactified on a CY 3-fold X is much more intricate. The exact kinetic terms for the hypermultiplet scalar is given by a negative QK manifold, but there are no non-renormalization theorems to prevent quantum corrections to it. There are several different sources of quantum corrections, perturbative and non-perturbative, some of which hard to control [122, 123]. Even the simplest class of corrections are as complicate as the GMN corrections to the HK metric in rigid $\mathcal{N} = 2$ SUSY [124]. At the very leading approximation, valid in the SUGRA approximation, we may start with Type IIB on the mirror manifold $\check{X}$ whose complex moduli space $\check{\mathcal{M}}$ is described by a special geometry which is protected by a non-renormalization theorem. We compactify this theory on a S^1 of radius R ; then $\check{\mathcal{M}}$ gets promoted to a quaternionic Kähler manifold $\check{Q}$ which in the SUGRA approximation (valid asymptotically as $R \rightarrow \infty$) is given in terms of the pre-potential $\check{F}$ of the special geometry $\check{\mathcal{M}}$ by the explicit formulae in [125] called the *c-map* between special Kähler manifolds and negative QK manifold [10]. $\check{Q}$ yields the *asymptotic geometry* of the quaternionic Kähler manifold Q for the original Type IIB on X in the limit in which the v.e.v. of the 4d light field in the original model which corresponds to the compactification radius R in the mirror model is sent to ∞ .

11.10 Global Aspects

Up to now we have reviewed the main *local* aspects of special Kähler geometry. In physics as well as in mathematics *global* aspects are essential. The special geometries which describe the moduli geometry of an actual family of CYs and/or the low-energy limit of a quantum consistent theory of gravity enjoy special global properties which are crucial for quantum consistency. For instance, we know that in these cases the global space $\mathcal{M}$ is quasi-projective: special Kähler spaces are *not* quasi-projective for “almost all” pre-potentials. So “almost all” formal 4d $\mathcal{N} = 2$ SUGRAS belong to the swampland. The global conditions yield a quite stringent selection rule on the physically sound special geometries.

When $\mathcal{M}$ is quasi-projective, the special geometry satisfies strong rigidity theorems, so that the full geometry (including the local metrics) is determined by the global behavior [62, 126]. We shall not dwell on the global issues, except for some very general comments. Let $\mathcal{M}$ be the actual moduli space, which is a quotient of the simply-connected special Kähler manifold $\tilde{\mathcal{M}}$ by a discrete group of isometries (which typically do not act freely—however $\mathcal{M}$ admits a smooth *finite* cover) and a quasi-projective manifold.

As a metric space $\mathcal{M}$ is almost never complete [61], so typically $\mathcal{M}$ is a “singular space”. These singularities do have a physical meaning: they lead to singularities of the low-energy supergravity, but they correspond to perfectly regular physical processes in the full superstring theory (which then manifest itself as the proper completion of the effective SUGRA at higher energies).

As already discussed, when we go along a non-trivial loop in $\mathcal{M}$ we come back with the electromagnetic duality frame rotated. This yields a group homomorphism the *monodromy representation*

$$\varrho: \pi_1(\mathcal{M}) \rightarrow Sp(2h^{2,1} + 1, \mathbb{Z}) \quad (11.337)$$

whose image is a discrete subgroup $\Gamma \subset Sp(2h^{2,1} + 1, \mathbb{Z})$, the *monodromy group*, which is necessarily infinite but may or may not be an arithmetic group.

The global aspects of the special geometry (and the existence of a UV completion of the associated supergravity) are controlled by Γ . For a readable review of the main properties of the monodromy group Γ see [52]. One important general result is

Theorem 11.12 (e.g. [127, 128]) *The Weil–Petersson volume of the actual complex moduli space $\mathcal{M}$ of a Calabi–Yau 3-fold is finite in fact a rational number.*

References

1. Griffiths, P.: Topics in Transcendental Algebraic Geometry. Princeton University Press (1984)
2. Carlson, J., Müller-Stach, S., Peters, C.: Period Mappings and Period Domains. Cambridge Studies in Advanced Mathematics, vol. 168, 2nd edn. Cambridge University Press, Cambridge (2017)
3. Voisin, C.: Hodge Theory and Complex Algebraic Geometry, vols. I & II. Cambridge University Press, Cambridge (2007)
4. Cecotti, S., Vafa, C.: Topological antitopological fusion. Nucl. Phys. B **367**, 359–461 (1991)
5. Cecotti, S., Vafa, C.: On classification of $N = 2$ supersymmetric theories. Commun. Math. Phys. **158**, 569–644 (1993). [arXiv:hep-th/9211097](https://arxiv.org/abs/hep-th/9211097)
6. Cecotti, S., Gaiotto, D., Vafa, C.: tt^* geometry in 3 and 4 dimensions. JHEP **05**, 055 (2014). [arXiv:1312.1008](https://arxiv.org/abs/1312.1008)
7. Cecotti, S.: Supersymmetric Field Theories. Geometric Structures and Dualities. Cambridge University Press, Cambridge (2015)
8. Freedman, D., van Proeyen, A.: Supergravity. CUP (2012)
9. Freed, D.S.: Special Kähler manifolds. Commun. Math. Phys. **302**, 31–52 (1999)
10. Cecotti, S., Ferrara, S., Girardello, L.: Geometry of type II superstrings and the moduli of superconformal field theories. Int. J. Mod. Phys. A **4**, 2475 (1989)

11. Cecotti, S.: $N = 2$ Supergravity, type IIB superstrings and algebraic geometry. *Commun. Math. Phys.* **131**, 517–536 (1990)
12. Strominger, A.: Special geometry. *Commun. Math. Phys.* **133**, 163–180 (1990)
13. Cecotti, S.: Swampland geometry and the gauge couplings. *JHEP* **09**, 136 (2021). [arXiv:2102.03205](https://arxiv.org/abs/2102.03205)
14. Candelas, P., Horowitz, G.T., Strominger, A., Witten, E.: Vacuum configurations for superstrings. *Nucl. Phys. B* **258**, 46–74 (1985)
15. A. Strominger, E. Witten New manifolds for superstring compactification, *Commun. Math. Phys.* **101** (1985) 341
16. Witten, E.: New issues in manifolds of $SU(3)$ holonomy. *Nucl. Phys. B* **283**, 79 (1986)
17. Kobayashi, S., Nomizu, K.: *Foundations of Differential Geometry*, vols. 1 and 2. Wiley, New York (1963)
18. Taubes, C.H.: *Differential Geometry. Bundles, Connections, Metrics and Curvature*. Oxford Graduate Texts in Mathematics, vol. 23. Oxford University Press, Oxford (2011)
19. Kobayashi, S.: *Transformation groups in Differential Geometry*. Classics in Mathematics. Springer, Berlin (1995)
20. Huybrechts, D.: *Complex Geometry*. Universitext, An Introduction. Springer, Berlin (2005)
21. Postnikov, M.M.: *Geometry VI. Riemannian Geometry*. Encyclopaedia Mathematical Sciences, vol. 91. Springer, Berlin (2001)
22. Kobayashi, S.: *Differential Geometry of Complex Vector Bundles*. Princeton University Press, Princeton (1987)
23. Besse, A.: *Einstein Manifolds*. Springer, Berlin (1987)
24. Joyce, D.D.: *Compact Manifolds with Special Holonomy*. Oxford Mathematical Monographs. Oxford University Press, Oxford (2000)
25. Joyce, D.D.: *Riemannian Holonomy Groups and Calibrated Geometry*. Oxford Graduate Texts in Mathematics, vol. 12. Oxford University Press, Oxford (2007)
26. Helgason, S.: *Differential Geometry. Lie Groups and Symmetric Spaces*. Academic, New York (1978)
27. Lerche, W., Vafa, C., Warner, N.P.: Chiral rings in $N = 2$ superconformal theories. *Nucl. Phys. B* **324**, 427 (1989)
28. Griffiths, P., Harris, J.: *Principles of Algebraic Geometry*. Wiley, New York (1978)
29. Chern, S.S.: *Complex Manifolds Without Potential Theory* (with an Appendix in the Geometry of Characteristic Classes). Springer, Berlin (1979)
30. Goldberger, S.I.: *Curvature and Homology*. Dover, Mineola (1982)
31. Cecotti, S.: $N = 2$ Landau-Ginzburg versus Calabi-Yau sigma models: nonperturbative aspects. *Int. J. Mod. Phys. A* **6**, 1749–1814 (1991)
32. Cecotti, S.: Geometry of $N = 2$ Landau-Ginzburg families. *Nucl. Phys. B* **355**, 755–776 (1991)
33. Barth, W., Hulek, K., Peter, C., van de Ven, A.: *Compact Complex Surfaces*. 2nd edn. *Erg. der Math. und ihrer Grenzgebiete, 3. Folge, Band 4*. Springer, Berlin (2004)
34. Joyce, D.D.: *Riemannian Holonomy Groups and Calibrated Geometry*. OUP (2007)
35. Beauville, A.: Variétés Kähleriennes dont la première classe de Chern est nulle. *J. Diff. Geom.* **18**, 755–782 (1983)
36. Oguiso, K., Sakurai, J.: Calabi-Yau threefolds of quotient type. *Asian J. Math.* **5** (2001). [arXiv:math/9909175](https://arxiv.org/abs/math/9909175)
37. Hashimoto, K., Kanazawa, A.: Calabi-Yau threefolds of type K (I): classification. [arXiv:1409.7601](https://arxiv.org/abs/1409.7601)
38. Hashimoto, K., Kanazawa, A.: Calabi-Yau threefolds of type K (II): mirror symmetry. [arXiv:1511.08778](https://arxiv.org/abs/1511.08778)
39. Hirzebruch, F.: *Topological Methods in Algebraic Geometry*. Classical in Mathematics. Springer, Berlin (1995)
40. Milnor, J.W., Stasheff, J.D.: *Characteristic Classes*. Princeton University Press, Princeton (1974)

41. Freedman, M.H.: The topology of four-dimensional manifolds. *J. Differ. Geometry* **17**, 357–453 (1982)
42. Freedman, M.H., Quinn, F.: (1990), *Topology of 4-Manifolds*. Princeton University Press, Princeton (1990)
43. Kodaira, K.: On the structure of compact complex analytic surfaces. I. *Amer. J. Math.* **86**, 751–798 (1964)
44. Huybrechts, D.: *Lectures on K3 Surfaces*. Cambridge University Press, Cambridge (2016)
45. Hübsch, T.: *Calabi-Yau Manifolds: A Bestiary for Physicists*. World Scientific, Singapore (1992)
46. Meyer, C.: *Modular Calabi–Yau Threefolds*. Fields Institute Monographs, vol. 22. AMS, Calgary (2005)
47. Bao, J., He, Y.H., Hirst, E., Pietromonaco, S.: *Lectures on the Calabi-Yau Landscape*. [arXiv:2001.01212](https://arxiv.org/abs/2001.01212)
48. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](https://arxiv.org/abs/1711.00864)
49. Kodaira, K.: *Complex Manifolds and Deformations of Complex Structures*. Springer, New York (1986)
50. Warner, F.W.: *Foundations of Differential Manifolds and Lie Groups*. Springer, Berlin (1983)
51. Yau, S.T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation. I. *Commun. Pure Appl. Math.* **31**, 339–411 (1978)
52. Peters, C.A.M., Steenbrink, J.H.M.: Monodromy of variations of Hodge structure. *Acta Appl. Math.* **75**, 183–194 (2003)
53. Hironaka, H.: Resolution of singularities of an algebraic variety of characteristic zero. *Ann. Math.* **79**, 109–326 (1964)
54. Lelong, P.: *Fonctions Plurisousharmoniques et Formes différentielles positives*. Gordon and Breach, Philadelphia (1969)
55. Grauert, H., Remmert, R.: *Coherent Analytic Sheaves*. Springer, Berlin (1984)
56. Neeman, A.: *Algebraic and Analytic Geometry*. Cambridge University Press, Cambridge (2007)
57. Chen, X., Liu, K., Shen, Y.: Global Torelli theorem for projective manifolds of Calabi-Yau type. [arXiv:1205.4207v3](https://arxiv.org/abs/1205.4207v3)
58. Liu, K., Shen, Y.: Hodge metric completion of the moduli space of Calabi-Yau manifolds. [arXiv:1305.0231](https://arxiv.org/abs/1305.0231)
59. Liu, K., Shen, Y.: From local Torelli to global Torelli. [arXiv:1512.08384](https://arxiv.org/abs/1512.08384)
60. Tian, G.: Smoothness of the universal deformation space of compact Calabi-Yau manifolds and Petersson-Weil metric. In: Yau, S.T. (ed.) *Mathematical Aspects of String Theory*, pp. 629–645. World Scientific, Singapore (1987)
61. Wang, C.-L.: On the incompleteness of the Weil-Petersson metric along degenerations of Calabi-Yau manifolds. *Math. Res. Lett.* **4**, 157–171 (1997)
62. Cecotti, S.: Special geometry and the swampland. *JHEP* **09**, 147 (2020). [arXiv:2004.06929](https://arxiv.org/abs/2004.06929)
63. Grisaru, M.T., Van de Ven, A.E.M., Zanon, D.: Four-loop β -function for the $N = 1$ and $N = 2$ supersymmetric non-linear sigma model in two dimensions. *Phys. Lett. B* **173**, 423–428 (1986)
64. Alvarez-Gaume, L., Freedman, D.Z.: Geometrical structure and ultraviolet finiteness in the supersymmetric sigma model. *Commun. Math. Phys.* **80**, 443 (1981)
65. Adler, S.L., Bardeen, W.A.: Absence of higher order corrections in the anomalous axial vector divergence. *Phys. Rev.* **182**, 1517 (1969)
66. Adler, S.L.: Anomalies to all orders. In: 't Hooft, G., (ed.), *Fifty Years of Yang-Mills Theory*, pp. 187–228. World Scientific, Singapore (2005). [arXiv:hep-th/0405040](https://arxiv.org/abs/hep-th/0405040)
67. Dine, M., Seiberg, N.: Nonrenormalization theorems in superstring theory. *Phys. Rev. Lett.* **57**, 2625–2628 (1986)
68. Zamolodchikov, A.B.: Irreversibility of the flux of the renormalization group in 2D field theory. *JETP Lett.* **43**, 730–732 (1986)
69. Gindikin, S.G., Pjateckii-Sapiro, I.I., Vinberg, E.B.: Classification and canonical realization of complex bounded homogeneous domains. In: *Transactions of the Moscow Mathematical Society for the Year 1963*, pp. 404–437. AMS, Calgary (1965)

70. Hori, K., Katz, S., Klemm, A., Pandharipande, R., Thomas, R., Vafa, C., Vakil, R., Zaslow, E.: *Mirror Symmetry*. Clay Mathematics Monographs, vol. 1. AMS, Clay Mathematical Institute (2003)
71. Skowronski, A.: Periodicity in representation theory of algebras. <https://webusers.imj-prg.fr/~bernhard.keller/ictp2006/lecturenotes/skowronski.pdf>
72. Keller, B.: Derived categories and tilting. In: Angeleri Hügel, L., Happel, D., Krause, H., (eds.), *Handbook of Tilting Theory*. London Mathematical Society Lecture Note Series, vol. 332. Cambridge University Press, Cambridge (2007)
73. Frölich, A., Taylor, M.J.: *Algebraic Number Theory*. Cambridge Studies in Advanced Mathematics, vol. 27. Cambridge University Press, Cambridge (1993)
74. Anspinwall, P.S., Bridgeland, T., Craw, A., Douglas, M.R., Gross, M., Kapustin, A., Moore, G.W., Segal, G., Szendrői, B., Wilson, P.M.H.: *Dirichlet Branes and Mirror Symmetry*. Clay Mathematical Monographs, vol. 4. AMS, Calgary (2009)
75. Dubrovin, B.: Geometry of 2-D Topological Field Theories. *Lectures Notes in Mathematics* **1620**, 120–348 (1996)
76. Witten, E.: Topological quantum field theory. *Commun. Math. Phys.* **117**, 411–449 (1988)
77. Witten, E.: Mirror manifolds and topological field theory. In: Yau, S.-T. (ed.) *Essays on Mirror Manifolds*. International Press, Vienna (1992)
78. Labastida, J., Marino, M.: *Topological Quantum Field Theory and Four Manifolds*. Springer, Berlin (2005)
79. Belavin, A.A., Polyakov, A.M., Zamolodchikov, A.B.: Infinite conformal symmetry in two-dimensional Quantum Field Theory. *Nucl. Phys. B* **241**, 333–380 (1984)
80. Di Francesco, P., Mathieu, P., Sánéchal, D.: *Conformal Field Theory*. Springer, Berlin (1997)
81. Boucher, W., Friedan, D., Kent, A.: Determinant formulae and unitarity for the $N = 2$ superconformal algebras in two dimensions or exact results on string compactification. *Phys. Lett. B* **172** 316
82. Zamolodchikov, A.B., Fadeev, V.A.: Disorder fields in two-dimensional conformal quantum field theory and $N = 2$ extended supersymmetry. *Sov. Phys. JETP* **63** 913
83. Qiu, Z.: Nonlocal current algebra and $N = 2$ superconformal field theory in two dimensions. *Phys. Lett. B* **188**, 207 (1987)
84. Gepner, D.: Spacetime supersymmetry in compactified string theory and superconformal models. *Nucl. Phys. B* **296**, 757 (1988)
85. Gepner, D.: Exactly solvable string compactifications on manifolds of $SU(N)$ holonomy. *Phys. Lett. B* **199**, 380–388 (1987)
86. Vafa, C.: String vacua and orbifoldized LG models. *Mod. Phys. Lett. A* **4** 1169
87. Morrison, D.R.: Compactifications of moduli spaces inspired by mirror symmetry. *Journées de Géométrie Algébrique*, Orsay 1992, ed. A. Beauville, Astérisque. [arXiv:alg-geom/9304007](https://arxiv.org/abs/hep-th/9311049)
88. Morrison, D.R.: Where is the large radius limit?. [arXiv:hep-th/9311049](https://arxiv.org/abs/hep-th/9311049)
89. Hori, K., Vafa, C.: Mirror symmetry. [arXiv:hep-th/0002222](https://arxiv.org/abs/hep-th/0002222)
90. Yau, S.-T. (ed.): *Essays on Mirror Manifolds*. International Press, Vienna (1992)
91. Candelas, P., de la Ossa, X.C., Green, P.S., Parkes, L.: A pair of Calabi-Yau manifolds as an exactly soluble superconformal theory. *Nucl. Phys. B* **359**, 21 (1991)
92. Bershadsky, M., Cecotti, S., Ooguri, H., Vafa, C.: Holomorphic anomalies in topological field theories. *Nucl. Phys. B* **405**, 279–304 (1993). [arXiv:hep-th/9302103](https://arxiv.org/abs/hep-th/9302103)
93. Bershadsky, M., Cecotti, S., Ooguri, H., Vafa, C.: Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes. *Commun. Math. Phys.* **165**, 311–428 (1994). [arXiv:hep-th/9309140](https://arxiv.org/abs/hep-th/9309140)
94. Katz, S.: *Enumerative geometry and string theory*. Student Mathematical Library. IAS/Park City Mathematical Subseries, vol. 32. AMS/IAS (2006)
95. Barbieri, R., Nanopoulos, D.V.: An exceptional model for grand unification. *Phys. Lett. B* **91**, 369–375 (1980)
96. Witten, E.: Superstring perturbation theory via super-Riemann surfaces: an overview. [arXiv:1304.2832](https://arxiv.org/abs/1304.2832)

97. Denef, F.: Les Houches lectures on constructing string vacua. *Les Houches* **87**, 483–610 (2008). [arXiv:0803.1194](#)
98. Tomasiello, A.: *Geometry of String Theory Compactifications*. Cambridge University Press, Cambridge (2022)
99. Shatashvili, S.L., Vafa, C.: Superstrings and manifold of exceptional holonomy. *Selecta Math.* **1**, 347 (1995). [arXiv:hep-th/9407025](#)
100. Acharya, B.S., Gukov, S.: M theory and singularities of exceptional holonomy manifolds. *Phys. Rept.* **392**, 121–189 (2004). [arXiv:hep-th/0409191](#)
101. de Wit, B., van Holten, J.W., van Proeyen, A.: Structure of $N = 2$ supergravity. *Nucl. Phys. B* **184**, 77 (1981)
102. de Wit, B., Lauwers, P.G., Philippe, R., Su, S., van Proeyen, A.: Gauge and matter fields coupled to $N = 2$ supergravity. *Phys. Lett. B* **134**, 37 (1984)
103. de Wit, B., Lauwers, P.G., van Proeyen, A.: Lagrangians of $N = 2$ supergravity - matter systems. *Nucl. Phys. B* **255**, 569 (1985)
104. de Roo, M., van Holten, J.W., de Wit, B., van Proeyen, A.: Chiral superfields in $N = 2$ supergravity. *Nucl. Phys. B* **173**, 175 (1980)
105. Bagger, J., Witten, E.: Matter couplings in $N = 2$ supergravity. *Nucl. Phys. B* **222**, 1 (1983)
106. Andrianopoli, L., Bertolini, M., Ceresole, A., D'Auria, R., Ferrara, S., Fré, P., Magri, T.: $N = 2$ supergravity and $N = 2$ super Yang-Mills theory on general scalar manifolds: symplectic covariance, gaugings and the momentum map. *J. Geom. Phys.* **23**, 111–189 (1997)
107. Lebrun, C., Salamon, S.: Strong rigidity of positive quaternionic Kähler manifolds. *Invent. Math.* **118**, 109–132 (1994)
108. Myers, S.B.: Riemannian manifolds with positive mean curvature. *Duke Math. J.* **8**, 401–404 (1941)
109. Deligne, P., de Griffiths, P.: *Séminaire Bourbaki Exp. 376*. Springer Lecture Notes in Mathematics, vol. 180. Springer, Berlin (1970)
110. Gaillard, M.K., Zumino, B.: Duality rotations for interacting fields. *Nucl. Phys. B* **193**, 221–244 (1981)
111. Cecotti, S., Ferrara, S., Girardello, L.: Hidden noncompact symmetries in string theory. *Nucl. Phys. B* **308**, 436–450 (1988)
112. Mumford, D.: *Abelian Varieties*. Reprinted by Hindustan Book Agency (2008)
113. Birkenhake, C., Lange, H.: *Complex Abelian Varieties*. A Series of Comprehensive Studies in Mathematics, vol. 302, 2nd edn. Springer, Berlin (2004)
114. Griffiths, P., Schmid, W.: Locally homogeneous complex manifolds. *Acta Math.* **123**, 145–166 (1969)
115. Bryant, R., Griffiths, P.: Some observations on the infinitesimal period relations for regular threefolds with trivial canonical bundle. In: Artin, M., Tate, J., (eds.), *Arithmetic and Geometry. Papers dedicated to I.R. Shafarevich*, vol. II, pp. 77–102. Birkhäuser (1983)
116. Bershadsky, M., Cecotti, S., Ooguri, H., Vafa, C.: Kodaira-Spencer theory of gravity and exact results for quantum string amplitudes. *Commun. Math. Phys.* **165**, 311–428 (1994). [arXiv:hep-th/9309140](#)
117. Boyer, C., Galicki, K.: *Sasakian Geometry*. Oxford Mathematical Monographs. Oxford University Press, Oxford (2008)
118. Cannas da Silva, A.: *Lectures on Symplectic Geometry*. Lectures Notes in Mathematics, vol. 1764. Springer, Berlin (2008)
119. Arnol'd, V.I.: The cohomology ring of the colored braid group. *Math. Notes* **5**, 138–140 (1969)
120. Guillemin, V., Sternberg, S.: *Symplectic Techniques in Physics*. Cambridge University Press, Cambridge (1990)
121. Witten, E., Bagger, J.: Quantization of Newton's constant in certain supergravity theories. *Phys. Lett. B* **115**, 202 (1982)
122. Alexandrov, S., Manschot, J., Persson, D., Pioline, B.: Quantum hypermultiplet moduli spaces in $N = 2$ string vacua: a review. *Proc. Symp. Pure Math.* **90**, 181–212 (2015). [arXiv:1304.0766](#)
123. Alexandrov, S., Persson, D., Pioline, B.: Wall-crossing, Rogers dilogarithm, and the QK/HK correspondence. *JHEP* **12**, 027 (2011). [arXiv:1110.0466](#)

- 124. Gaiotto, D., Moore, G.W., Neitzke, A.: Four-dimensional wall-crossing via three-dimensional field theory. *Commun. Math. Phys.* **299**, 163–224 (2010). [arXiv:0807.4723](#)
- 125. Ferrara, S., Sabharwal, S.: Quaternionic manifolds for type II superstring vacua of Calabi-Yau spaces. *Nucl. Phys. B* **332**, 317–332 (1990)
- 126. Cecotti, S.: Moduli spaces of Calabi-Yau d -folds as gravitational-chiral instantons. *JHEP* **12**, 008 (2020). [arXiv:2007.09992](#)
- 127. Lu, Z., Sun, X.: Weil-Petersson geometry on moduli space of polarized Calabi-Yau manifolds. *J. Inst. Math. Jussieu* **3**, 185–229 (2004)
- 128. Lu, Z., Sun, X.: On the Weil-Petersson volume and the first Chern class of the moduli space of Calabi-Yau manifolds. *Commun. Math. Phys.* **261** 297 (2006). [arXiv:math.dg/0510021](#)

Part IV

Superstrings Beyond Weak Coupling

In this Part, we discuss non-perturbative phenomena of the supersymmetric strings with particular reference to their BPS sectors. The key tool is SUSY nonrenormalization properties.

Part IV consists of three chapters.

In Chap. 12, we discuss T -duality, D -branes, and orientifolds in the context of supersymmetric string theories. Here we have many new phenomena and interesting questions. The main focus is on the extended BPS objects and their physical implications.

In Chap. 13, we analyze the supersymmetric string theories at strong coupling using the BPS objects to control their non-perturbative physics.

Chapter 14 contains additional topics and further directions mainly meant as invitations to more advanced materials and physical theories.

Chapter 12

Superstring D-Branes



Abstract We study T -duality, D-branes, and orbifolds in the Type II and Type I superstring theories where they present new interesting phenomena. We compute the tension and quantum numbers of these extended objects, and determine the effective action of the d.o.f. living on the branes. We address the fundamental issue of the existence of multi-object and bound states of D-branes and fundamental strings which are BPS, i.e. preserve some supersymmetry. We study the physics of these BPS objects from many complementary viewpoints: the spacetime perspective, the world-sheet theory, and the QFT on the branes. We describe the geometry of the BPS objects and outline the relation with YM instanton and their ADHM construction

12.1 T -Duality in Type II Strings

In Chap. 6, we studied T -duality and its consequences in the context of the bosonic string. T -duality of supersymmetric strings is a much deeper story with surprising dynamical implications. We consider first closed oriented Type II theories. In either IIA or IIB, we compactify a single coordinate, X^9 , on a circle of radius R and take the limit $R \rightarrow 0$. As in Chap. 6, this is equivalent to the $R' \rightarrow \infty$ limit in the dual coordinate, $X^{9'}$, whose right-hand part gets reflected

$$X_R^{9'}(\bar{z}) = -X_R^9(\bar{z}) \quad (12.1)$$

just as in the bosonic string. To preserve 2d superconformal invariance, we must reflect also its fermionic superpartner

$$\tilde{\psi}^{9'}(\bar{z}) = -\tilde{\psi}^9(\bar{z}). \quad (12.2)$$

Equation (12.2) implies that the chirality of the right-moving R-sector ground state is reversed: the raising and lowering Clifford operators $\Gamma^\pm \equiv \tilde{\psi}_0^8 \pm i\tilde{\psi}_0^9$ get interchanged. In simpler terms: T -duality is a spacetime parity transformation acting only on the right-movers, and so it reverses the relative chiralities of the right- and left-moving ground states. If we start with IIA and take the radius to be small, we

obtain IIB at large radius, and vice versa. The same holds if we T -dualize (i.e. make the above one-side reflection) on any *odd* number of directions, while T -dualizing an *even* number of directions gives back the original Type II theory. As for the two 10d heterotic theories in Sect. 7.7.1, we conclude that *the two inequivalent 10d Type II superstrings are two different limits of a single connected space of compactified theories.*

IIA and IIB have different sets of RR fields, and T -duality must interchange the two sets. To preserve the Lorentz $SO(9, 1)$ current algebra (i.e. the OPE between $\tilde{\psi}^\mu$ and the spin field, Eqs. (2.490)–(2.493)) T -duality along the 9-axis must act as¹

$$S_\alpha(z)' = S_\alpha(z), \quad \tilde{S}_\alpha(\bar{z}) = (\beta^9)_\alpha^\beta \tilde{S}_\beta(\bar{z}), \quad (12.3)$$

where β^9 is the matrix which implements the parity transformation (reflection of the 9th axis) on the spinors. By its very definition, β^9 anticommutes with Γ^9 and commutes with the other Γ^μ , so, up to an overall phase,

$$\beta^9 = \Gamma^9 \Gamma \quad (12.4)$$

where $\Gamma \equiv \Gamma_{11}$ is the chirality matrix in 10 dimensions. We chose the overall phase in (12.4) so that β^9 is real in a Majorana representation, i.e. so that β^9 maps MW spinors $\tilde{S}$ into MW spinors $\beta^9 \tilde{S}$ of opposite chirality. T -duality acts on the RR vertices,²

$$S^t C \Gamma^{\mu_1 \cdots \mu_{p+1}} \tilde{S} e^{-(\phi + \tilde{\phi})/2} e^{ikX}, \quad (12.5)$$

by multiplying the product of Γ matrices by $\Gamma^9 \Gamma$ on the right

$$C \Gamma^{\mu_1 \cdots \mu_{p+1}} \rightarrow C \Gamma^{\mu_1 \cdots \mu_{p+1}} \Gamma^9 \Gamma, \quad (12.6)$$

Γ gives ± 1 because the right-moving spin fields $\tilde{S}$ have a definite chirality by GSO projection. Then Γ^9 adds a 9 to the index set $\mu_1 \cdots \mu_{p+1}$ if none is present, or removes the 9 if it is present since $(\Gamma^9)^2 = 1$ (for a more systematic treatment; cf. **BOX 12.1**). Thus T -duality acts on the RR field strengths and potentials by adding or subtracting the index of the dualized direction. Therefore, *up to sign*, T -duality maps IIA RR fields into IIB RR fields according to the rule

$$\begin{aligned} C_9^{(1)} &\rightarrow C^{(0)}, \\ C_\mu^{(1)}, C_{\mu\nu 9}^{(3)} &\rightarrow C_{\mu 9}^{(2)}, C_{\mu\nu}^{(2)} \\ C_{\mu\nu\lambda}^{(3)} &\rightarrow C_{\mu\nu\lambda 9}^{(4)}, \end{aligned} \quad (12.7)$$

¹ We write the spin fields as Dirac spinors rather than Weyl spinors as elsewhere in this book.

² Recall from Sect. 3.7.3 that the $(-\frac{1}{2}, -\frac{1}{2})$ -picture RR vertices are linear in the field strengths $F^{(p+1)} = dC^{(p)}$ of the RR gauge p -form fields $C^{(p)}$ rather than in the gauge fields $C^{(p)}$ themselves.

BOX 12.1 - More on the product in Clifford algebras

We write $\text{Cl}_+(n)$ for the Clifford algebra generated by Γ^a , $a = 1, \dots, n$ with the relation

$$\Gamma^a \Gamma^b + \Gamma^b \Gamma^a = 2 \delta^{ab}.$$

The elements of the basis of $\text{Cl}_+(n)$ are in one-to-one correspondence with the subsets I of the set $[n] \equiv \{1, 2, \dots, n\}$. The element Γ_I associated with the subset $\{i_1 < i_2 < \dots < i_m\} \subset [n]$ is $\Gamma^{i_1} \Gamma^{i_2} \dots \Gamma^{i_m}$. The multiplication table in $\text{Cl}_+(n)$ is

$$\Gamma_I \Gamma_J = (-1)^{\tau_+(I, J)} \Gamma_{I \Delta J},$$

where $I \Delta J \equiv (I \cup J) \setminus (I \cap J)$ is the *symmetric difference* of the two sets I and J and

$$\tau_+(I, J) = \#\{(i, j) \in I \times J \mid i > j\}$$

where $\mu, \nu, \lambda \neq 9$. We could go on, getting $C_{\mu\nu\lambda\omega}$ from $C_{\mu\nu\lambda\omega 9}$ and so on, but these higher-degree form-fields are not independent, but rather they are dual descriptions of the same degrees of freedom: indeed the massless vertex (12.5), seen as a $(p+1)$ -form, is either self-dual or anti-self-dual, given that

$$\Gamma^{\mu_1 \dots \mu_{p+1}} = \frac{1}{(9-p)!} \epsilon^{\mu_1 \dots \mu_{p+1} \mu_{p+2} \dots \mu_{10}} \Gamma_{\mu_{p+1} \dots \mu_{10}} \Gamma \quad (12.8)$$

while $\Gamma \tilde{S} = \pm \tilde{S}$ by GSO projection. Hence two dual RR field strengths

$$F^{(p+1)} = dC^{(p)} \quad \text{and} \quad F^{9-p} \equiv *F^{(p+1)} = dC^{(8-p)} \quad (12.9)$$

produce the *same* BRST-invariant 2d vertex operator and hence (by the CFT state-operator correspondence) yield the *same* physical state of the superstring.

For T -duality along several axes we replace β^9 with

$$\beta^\perp \stackrel{\text{def}}{=} \prod_m \beta^m \quad (12.10)$$

where $\beta^m = \Gamma^m \Gamma$ and the product is over all T -dualized axes. We need to keep track of some important signs: since $\beta^m \beta^n = -\beta^n \beta^m$, for $m \neq n$, T -dualities in different directions do not commute but differ by a sign in the right-moving R sector. We write this as

$$\beta^m \beta^n = (-1)^{\tilde{\mathbf{F}}} \beta^n \beta^m, \quad (12.11)$$

where $\tilde{\mathbf{F}}$ is the spacetime Fermi number on the right-moving state of the string: $(-1)^{\tilde{\mathbf{F}}}$ is a symmetry which flips the sign of all right-moving R states. With our choice of

phase, $(\beta^m)^2 = -1$, so acting twice with the same T -duality produces $(-1)^{\tilde{F}}$ acting on the original vertices.

12.2 T -Duality of Type I Strings: SUSY D-Branes

Taking the T -dual of the open and unoriented Type I $SO(32)$ string leads to D-branes and orientifold planes by the same arguments as in the bosonic string; cf. Chap. 6. In particular, the T -dual of Type I compactified on a circle S^1 is a configuration of 16 D8-branes placed between two orientifold hyperplanes. The positions of the D8-branes in the dual description correspond to the eigenvalues of the Wilson line ($\equiv$ gauge holonomy) on S^1 of the original Type I.

Let us consider the bulk physics of the T -dual theory, obtained by taking $R \rightarrow 0$ and focusing on a region of the decompactified dual spacetime that is far away from the fixed planes and the D-branes. Just as in the bosonic string, the local physics is that of a closed oriented superstring theory: *closed* because the open strings live far away from the D-branes; *oriented* since the orientation projection Ω relates the state of a string to that of its image on the other side of the fixed plane, but does not give local conditions on the string states. The local physics must be described by a Type II theory, hence *two* massless gravitini propagate in the bulk, so, to avoid inconsistencies,³ all closed string processes must be invariant under 32 supersymmetries. Type I has equal left- and right-chiralities; taking the T -dual on one direction makes them *opposite*: the local physics in the bulk is of Type IIA. Taking the T -dual on an odd number of dimensions yields the same theory; taking the T -dual on an even number of coordinates gives Type IIB in the bulk.

Next we compactify Type I on $(S^1)^{9-p}$ and take all radii to 0, while concentrating on the vicinity of one D-brane in the T -dual theory, adjusting the Wilson lines so that the fixed plane and the other D-branes move away in the decompactified T -dual spacetime. We remain with 10d Type II with an isolate Dp -brane where p is even for IIA and odd for IIB.

Dp -Branes as BPS Objects

We consider a process where closed strings scatter with a D-brane. To compute the amplitude we must use world-sheets Σ having a boundary on the brane. The open string boundary conditions of the original Type I preserve 10d $\mathcal{N} = 1$ SUSY: the left-moving world-sheet current of spacetime supersymmetry $e^{-\phi/2} S_\alpha(z)$ flows into the boundary and the right-moving current $e^{-\tilde{\phi}/2} \tilde{S}_\alpha(\bar{z})$ flows out, and only the total supercharge

$$Q_\alpha + \tilde{Q}_\alpha \equiv \oint \frac{dz}{2\pi i} e^{-\phi/2} S_\alpha(z) + \oint \frac{d\bar{z}}{-2\pi i} e^{-\tilde{\phi}/2} \tilde{S}_\alpha(\bar{z}) \quad (12.12)$$

³ We recall again that non-free massless spin- $\frac{3}{2}$ particles propagate consistently only if the full theory is supersymmetric [1].

is conserved. Under T -duality the conserved supercharges become

$$Q'_\alpha + (\beta^\perp \tilde{Q}')_\alpha, \quad (12.13)$$

and again the scattering amplitudes of closed strings with the D-brane are invariant under 16 supersymmetries of the form (12.13).

We compare the situation with the conservation of momentum. There is a non-zero amplitude for a closed string to reflect backwards from the D-brane; this process does not conserve momentum in the direction orthogonal to the D-brane. From the world-sheet perspective, this is due to the Dirichlet b.c. which explicitly breaks translational invariance. From the spacetime point of view, however, this symmetry breaking is *spontaneous*: we are expanding around a configuration with a D-brane placed in some definite position, but there are states of the same energy with the D-brane translated in any place.⁴ In a spontaneously broken symmetry the apparent violation of the conservation law is related by the Ward identities to the amplitude to emit a soft Goldstone boson. For the D-brane, as for any extended object, the Goldstone bosons of the broken translations are the collective coordinates for its motion. Let us see this in detail: the non-conservation of momentum is measured by the integral of the corresponding Noether current over the world-sheet boundary

$$\frac{1}{2\pi\alpha'} \int_{\partial\Sigma} ds \partial_n X^{9'}, \quad (12.14)$$

which (up to normalization) is just the 0-picture vertex operator for the collective coordinate, with zero momentum in the Neumann directions. By the SUSY algebra, the same holds for the supercharges: the D-brane breaks *spontaneously* 16 of the 32 spacetime supersymmetries, the ones which, from the 2d perspective, are explicitly broken by the open string boundary conditions. The integrals

$$\int_{\partial\Sigma} ds S'_\alpha e^{-\phi/2} = - \int_{\partial\Sigma} ds (\beta^9 \tilde{S}')_\alpha e^{-\tilde{\phi}/2} \quad (12.15)$$

which measure the breaking of SUSY are the vertices of the Fermi open string states. These states form a *goldstino*, the Goldstone particle of spontaneous breaking of supersymmetry. The goldstino propagates only along the brane since in the bulk all supersymmetries are unbroken. The fact that D-branes break SUSY is not surprising: the unique state invariant for *all* supersymmetries is the vacuum. What is remarkable is that the D-brane leaves *half* SUSY unbroken: *the D-brane is a BPS state*. The

⁴ There is a third point of view: the effective $(p+1)$ -dimensional field theory of the light d.o.f. propagating on the world-volume of the D p -brane. In this case the translational symmetry is broken *spontaneously* and the massless scalar fields—which represent orthogonal motions of the brane—are the corresponding Goldstone particles in this $d = p+1$ QFT. However when $p=1$ we have a subtlety: we cannot have spontaneous symmetry breaking by the Mermin–Wagner–Coleman theorem [2]. Thus the D1 brane enjoys a somewhat different status. We shall understand why when studying superstring theory at strong coupling.

argument holds for any number of T -dualized coordinates, and so for all Dp -branes. The unbroken supersymmetries are

$$Q_\alpha + (\beta^\perp \tilde{Q}')_\alpha, \quad (12.16)$$

with, as before, $\beta^\perp = \prod_m \beta^m$ where the product is over the directions perpendicular to the D-brane. The superstring D-branes, being BPS, are absolutely *stable*. This should be contrasted with the D-branes of the bosonic string which were not stable: indeed they had a tachyonic mode.

Massless Fields Propagating on the D-Brane World-Volume.

The low energy d.o.f. on an isolated D8-brane are the massless open string states

$$\psi_{-1/2}^\mu |k\rangle_{\text{NS}}, \quad \psi_{-1/2}^9 |k\rangle_{\text{NS}}, \quad |\alpha, k\rangle_{\text{R}}. \quad (12.17)$$

Just as in the bosonic theory Sect. 6.9, the bosonic states are a gauge field A_a living on the D-brane and the collective coordinates X^9 for the D-brane which describe its motions in normal directions. The fermionic states $|\alpha, k\rangle_{\text{R}}$ are the superpartners of these, that is, a *photino* with 16 components (8 on-shell). As in the bosonic string, the momentum k^μ of these states is constrained to be tangent to the brane, so these states propagate only on the brane world-volume. The massless d.o.f. on the world-volume of an isolated D8-brane form one $U(1)$ vector supermultiplet of $\mathcal{N} = 1$ SUSY in $d = 9$. More generally, the light d.o.f. living on an isolated Dp -brane is a $U(1)$ vector supermultiplet of⁵ $\mathcal{N} = 16/N(d)$ SUSY in $d = p + 1$.

Going through the same analysis as in Chap. 6 for the bosonic string, and using that the D-brane configuration preserves 16 supercharges, we conclude that the massless d.o.f. propagating along the world-volume of a Dp -brane are

- for a stack of n parallel Dp -branes away from orientifold planes a $U(n)$ SYM supermultiplet in $d = p + 1$ invariant under 16 supercharges;
- for a stack of n Dp -branes on an orientifold plane a $SO(2n)$ SYM supermultiplet in $d = p + 1$ invariant under 16 supercharges.

Couplings to RR Gauge Potentials

By the extended SUSY algebra [3, 4], massive BPS states must carry conserved charges. Indeed, their mass saturates a BPS bound of the form $m \geq |Z|$ where Z is a (complex) conserved charge which can be written as a linear combination of the conserved quantities $\{Q, \tilde{Q}\}$ and P_μ . For extended objects, like branes, both mass and charge are extensive quantities, and we must replace m and Z by the corresponding quantities per *unit volume*.

In the present context, there is a natural set of charges with the correct Lorentz structure to be carried by Dp -branes, i.e. the RR-charges associated with the RR gauge-form fields $C^{(k)}$. A p -brane has a natural coupling to a $(p + 1)$ -form field $C^{(p+1)}$

⁵ Here $N(d)$ is the function defined in Eq. (8.23).

$$\mu_p \int_M C^{(p+1)}, \quad (12.18)$$

where the integral is over the $(p + 1)$ -dimensional world-volume M which describes the world-story of the Dp -brane. The interaction (12.18) is the obvious generalization of the electric coupling of the Abelian gauge field $A^{(1)}$ to a charged point particle.⁶ The overall constant μ_p measures the charge of the D-brane per unit volume.

By a chain of *T*-dualities we can get Type IIA with a Dp -brane for any *even* p . Thus in Type IIA we need 1-, 3-, 5-, 7-, and 9-form potentials. Between the propagating massless RR d.o.f. of the IIA theory there are a 1-form and a 3-form, while the 5-form and 7-form give equivalent dual description of the same physics (cf. discussion around Eqs. (12.8), (12.9)). The non-propagating 9-form potential was found in the careful BRST treatment at the end of Sect. 3.7.3, and further discussed in the context of massive IIA SUGRA (see Sect. 8.4 and **BOX 3.3**). Although $C^{(9)}$ does not describe propagating states (its momentum is frozen to 0), the very existence of D8-branes shows that $C^{(9)}$ is part of the IIA theory.⁷ This is in full agreement with our BRST analysis in Chap. 3.

It follows from our discussion around Eqs. (12.8), (12.9) that a Dirichlet p -brane and a Dirichlet $(6 - p)$ -brane are, respectively, electric and magnetic sources for the *same* gauge field $C^{(p+1)}$. For example, the free field equation of motion and the Bianchi identity for a 2-form field strength

$$d * F^{(2)} = dF^{(2)} = 0, \quad (12.19)$$

are symmetric between $F^{(2)}$ and its dual 8-form

$$F^{(8)} \equiv *F^{(2)} \quad (12.20)$$

and they can be written in terms of either a 1-form or a 7-form gauge potential

$$F^{(2)} = dC^{(1)}, \quad d * dC^{(1)} = 0, \quad (12.21)$$

$$*F^{(2)} = (*F)^{(8)} = dC^{(7)}, \quad d * dC^{(7)} = 0. \quad (12.22)$$

At the location of an *electric* source, which would be a D0-brane for $C^{(1)}$ and a D6-brane for $C^{(7)}$, the field equation has a source term

$$d * dC^{(p+1)} = \text{source with support on the } Dp\text{-brane.} \quad (12.23)$$

At a *magnetic* source, which is a D6-brane for $C^{(1)}$ and a D0-brane for $C^{(7)}$, the Bianchi identity breaks down, and the potential cannot be globally defined: one must

⁶ Compare (12.18) with the electric coupling $e \int_\ell A^{(1)}$ of a *charged* point particle with a vector potential $A^{(1)}$ (ℓ is the particle world-line).

⁷ Since the gauge field $C^{(9)}$ is an unusual field, one expects its source, the D8-brane, to be a peculiar object with a different physics. We shall confirm this when discussing branes at strong coupling.

introduce Dirac “strings” (actually Dirac $(p + 1)$ -branes) or use different potentials in the coordinate patches related by non-trivial gauge transformations in the overlaps.⁸

For Type IIB we need 2-, 4-, 6-, and 10-form potentials. The first three arise in either the electric or magnetic description of the propagating RR states. The existence of the 10-form potential $C^{(10)}$ was deduced from the study of Type I divergences (Sect. 5.6) and from the Dirac–Kähler form of the BRST condition (BOX 3.3). In Sect. 5.6, we showed that the coupling has the form

$$\mu_9 \int C^{(10)} \quad (12.24)$$

where the integral runs over all spacetime. This fits with the idea that the proper interpretation of the Chan–Paton d.o.f. in the fully Neumann sector (i.e. in the original Type I) is to label the several 9-branes on which each endpoint of the string may lay. The 9-branes fill all spacetime so do not break 10d Poincaré invariance.

We stress that all the other RR couplings (12.18) follow from the already known one, Eq. (12.24), by a chain of T -dualities, since each time we T -dualize an additional coordinate the dimension of the p -brane goes down by one and the RR- form loses one index; cf. (12.7). The chain of dualities also yields a recursion relation on the coupling constants μ_p (see below).

The IIB theory also has a 0-form RR potential $C^{(0)} \equiv \text{Re } \tau$, the RR scalar. This scalar should couple to a (-1) -brane. There is an obvious interpretation for a (-1) -brane: it is defined by Dirichlet boundary conditions in all directions, time as well space, so its world-sheet is zero-dimensional, hence a point, and the integral (12.18) reduces to the value of $C^{(0)}$ at this point. An object which is localized in time as well as in space is an *instanton*. In the Euclidean path integral formalism, instantons describe classically-forbidden quantum tunneling processes. Unitarity requires these processes to be present in string theory, so the existence of a D-instanton should not be a surprise. A direct connection between the D (-1) -brane and the YM instanton will be made explicit in Sect. 12.8.

In Sect. 12.4 we shall give a direct proof that D-branes carry RR charges, that is,

$$\mu_p \neq 0, \quad (12.25)$$

although this should be already obvious from the fact that μ_9 is non-zero (we computed it in Sect. 5.6 as a RR tadpole) and all other μ_p ’s follow from it by T -duality. Before going to the proof, we discuss some physical implications of this result.

Note 12.1 We saw in Sect. 3.7.3 that *perturbative* string states do not carry RR charges, but the computation in Sect. 12.4 will show that superstring theory has sources for every RR gauge fields. For the NS-NS gauge field $B_{\mu\nu}$ all ordinary string states are neutral, but it couples *electrically* to the winding number of strings and we have seen in Sect. 8.8 (cf. Eqs. (8.135), (8.137)) that there is also a dual 5-brane

⁸ Same story, *mutatis mutandis*, as for the Dirac monopole in the 4d electrodynamics, see, e.g. [5].

magnetic source for $B_{\mu\nu}$ (the NS5-brane), thus vindicating our claim that in string theory *all gauge fields have a full set of electric and magnetic source objects*.

RR Charge of Orientifold Planes

In Sect. 5.6 the divergence of Type I for gauge groups other than $SO(32)$ arose from the inconsistency of the RR 10-form field equation, see Eq. (5.138). This divergence is unaffected by toroidal compactification, and again cancels only for $G = SO(32)$.

The spacetime interpretation of the divergence in the T -dual picture with D-branes is again an inconsistency in the RR field equations. Gauge field flux-lines emerge from each D-brane, orthogonal to the non-compact directions, and these flux-lines should end somewhere. Moreover, all D-branes must have charges of the *same sign*: the full set of D-branes is a BPS state, being T -dual to the Type I theory (invariant under 16 SUSY), and the total mass is *linear* in the total charge for a BPS state.⁹ From Sect. 5.6 we know that for $G = SO(32)$ the disk tadpole is canceled by the contribution of the unoriented cross-cap. In the T -dual spacetime the cross-cap must be localized at one of the orientifold planes, since the strings are oriented in the bulk. We conclude that the orientifold planes are *sinks* for the RR charge. If we T -dualize k coordinates, we get 2^k orientifold planes but still 16 D-branes, so that the RR charge of an orientifold plane should be

$$-2^{4-k} \quad (12.26)$$

times that of a D-brane of the same dimension.

Note 12.2 The toroidal compactifications we are considering are the ordinary ones in which the topology of the gauge bundle is “trivial”. There is a subtler possibility pointed out in particular by Witten [6]: *toroidal compactifications without vector structure*. This leads to a different kind of orientifold planes with opposite charge $+2^{4-k}$ and gauge group $Sp(k)$ instead of $SO(n)$ which we may think of as arising from the other sign in the Ω -projection of CP indices (cf. Sect. 3.9). One speaks of $\pm Op$ orientifold planes where the sign is the sign of its charge. For more advanced material on orientifolds in Type II see [7–9].

12.3 Relations Between Superstring Theories

Starting from the toroidally compactified Type I theory we can reach both $d = 10$ Type II theories. Simply take an odd or even number of radii to zero, while moving the D-branes and fixed planes off to infinity as the dual space expands. Thus, just as for the two heterotic theories, the two Type II and Type I should be regarded as different limits of a *single* theory. The theory has many other states: we may take the limit $R' \rightarrow \infty$ while keeping some of the D-branes in fixed positions, so that we obtain the 10d theory with D-branes. By itself, T -duality produces only configurations of parallel D-branes all of the same dimension, but the D-branes are dynamical objects

⁹ Cf. the general discussion of BPS objects in Sect. 8.2.

whose position and shape evolve in time, and we may continuously deform their configurations. We can also construct states with p -branes of different dimensions by the following process: we start with two D1-branes in type IIB. The first D1 is, say, stretched along the 1-axis and we rotate the second one along the 2-axis. T -duality along the 2-axis reverses Dirichlet and Neumann b.c. and turns the first D-string into a D2-brane extended in the 1–2 plane and the second one into a D0-brane. Thus a D2 and a D0 can coexist in Type IIA. T -duality leads to states with at most 16 D-branes, but this is due to the RR flux conservation in the compact space. In a non-compact space the RR flux can flow to infinity, and any number of D-branes is allowed (except for D9's which fill spacetime, leading to a dangerous tadpole instead of a flux). All these statements are obvious from the world-sheet perspective, where D-branes are just boundary conditions and we can combine them in various ways.

Thus, starting from Type I, we can reach Type IIA configurations with any collection of Dp -branes with *even* p 's or Type IIB ones with any of Dp -branes with *odd* p 's. In conclusion: we have a single theory which has a IIA sector with no D-brane, a IIB sector with no D-branes (T -dual to the first one), a Type I sector with 16 D9-brane and an orientifold 9-plane, and infinitely many sectors with general arrangements of D-branes and orientifold planes all of the same dimension mod 2.

SUSY Algebra

We write the SUSY algebra for IIA/IIB in presence of strings and Dp -branes (p even/odd, respectively)

$$\{Q_\alpha, \bar{Q}_\beta\} = -2[P_M + (2\pi\alpha')^{-1} Q_M^{NS}] \Gamma_{\alpha\beta}^M - 2 \frac{\tau_{NS}}{5!} Q_{M_1 \dots M_5}^{NS} \Gamma_{\alpha\beta}^{M_1 \dots M_5} \quad (12.27)$$

$$\{\bar{Q}_\alpha, \bar{Q}_\beta\} = -2[P_M - (2\pi\alpha')^{-1} Q_M^{NS}] \Gamma_{\alpha\beta}^M + 2 \frac{\tau_{NS}}{5!} Q_{M_1 \dots M_5}^{NS} \Gamma_{\alpha\beta}^{M_1 \dots M_5} \quad (12.28)$$

$$\{Q_\alpha, \bar{Q}_\beta\} = -2 \sum_p \frac{\tau_p}{p!} Q_{M_1 M_2 \dots M_p}^R (\beta^{M_1} \beta^{M_2} \dots \beta^{M_p})_{\alpha\beta}. \quad (12.29)$$

Here:

- the spacetime supercharges $Q_\alpha, \bar{Q}_\alpha$ act respectively on the left- and right-movers
- as in Sect. 7.8, the anticommutator of two right-moving supercharges contains the winding charge Q^{NS} which couples (electrically) to the NS-NS 2-form B . The same holds for the left-moving supercharges. Q^{NS} is the charge of the *fundamental string*. We call it the F -string to distinguish it from the D1-brane (D -string);
- dually $Q_{M_1 \dots M_5}$ is the magnetic charge which couples to the 6-form magnetic dual of B . We found in Sect. 8.8 that there is a BPS object—the *NS5 magnetic five-brane soliton*—carrying this magnetic charge;
- $Q_{M_1 M_2 \dots M_p}^R$ are the RR Dp charges;
- all charges are normalized to 1 per unit volume of the corresponding extended object, so the brane tensions τ_{NS}, τ_p and the string tension $(2\pi\alpha')^{-1}$ appear explicitly in the SUSY algebra;

- the anticommutator of a left- and a right- supersymmetry is fixed by imposing that D-branes are BPS object. The sum in p runs over the even (resp. odd) values for IIA (resp. IIB). τ_p is the D p -brane tension.

To check that the algebra is correct, we consider a state which contains a single D p -brane; its non-zero charge is $Q_{\mu_1 \dots \mu_p}^R = 1$ where the indices $\mu_1, \dots, \mu_p$ run over the directions tangent to the D p -brane. One has

$$\beta^{\mu_1} \dots \beta^{\mu_p} = \pm \beta^\perp \Gamma^0. \quad (12.30)$$

The sign may be absorbed in the conventions. It then follows that the anticommutator of $Q + \beta^\perp \tilde{Q}$ vanishes in this state, as required by the BPS property.

Note 12.3 The heterotic world-sheet theory has no boundary conditions which preserve the world-sheet gauge symmetries, hence no objects akin to D-branes exist.

12.4 D-Brane Tensions and RR Charges

There is no net force between objects which are all invariant under a common (non-empty) set of supercharges: the multi-object state is still supersymmetric, saturating the appropriate BPS bound, and so its energy is determined by its charges and is independent of the separations. This applies in particular to BPS objects at relative rest carrying the same central charge. Indeed, the T -duality construction shows that the unbroken supersymmetry of a set of static parallel D p branes is the same one as for a single D p -brane.

The vanishing of the force between parallel D p branes results from a perfect cancelation between the attraction due to the graviton and dilaton and the repulsion due to the exchange of RR gauge particles. We can calculate these forces explicitly from the usual cylinder vacuum amplitude as we did in Sect. 6.9.1 for the D-branes of the bosonic string, see Fig. 6.3. At the full superstring level, the RR exchange force is *minus* the one for the NS-NS exchange. This can be understood in two ways:

- (1) the configuration of two parallel D p -branes *at relative rest* is BPS for all separation of the two branes, so that their energy (equal to $2|Z|$ in the common rest frame) is independent of their separation, hence there is no net force between them;
- (2) the RR plus NS-NS amplitude computes the Witten index of the *refermionized* 2d world-sheet SCFT quantized on the cylinder with boundary conditions which preserve the 2d supercharge $G_0 + \tilde{G}_0$.

The exchange of light NS-NS closed string was isolated in Sect. 5.6 when we computed the one-loop cylinder amplitude for oriented open superstrings, see Eq. (5.127) that we reproduce here for the convenience of the reader (we set $s = \pi \ell \equiv \pi/t$, where t is the length of the circle in the cylinder $[0, \pi] \times S^1$, and reintroduce α')

$$Z_0 = \frac{i N^2}{8\pi(8\alpha'\pi^2)^5} \int_0^\infty ds [16 + O(e^{-2s})]. \quad (12.31)$$

In the T -dual language the cylinder amplitude (12.31) is interpreted as the tree-level exchange of NS-NS closed superstring states between the two Dp -branes associated with the Dirichlet b.c. at the two boundaries, see Fig. 6.3. As in the last paragraph of Sect. 6.9.1, we have to omit the factor N^2 and the integration over the frozen momentum in the $(9-p)$ Dirichlet directions, and also shift the open-string Hamiltonian by the tension term (6.298) of a string suspended between parallel D-branes separated by y^i in the i -th normal direction. Including an extra factor 2 from the two orientations of the string, and omitting the factor $1/2$ which in Sect. 5.6 came from the orientation projection (recall that now the strings are *locally* oriented), for large separations $|y^i| \gg \sqrt{\alpha'}$ we get the NS-NS amplitude [10, 11]

$$\begin{aligned} \mathcal{A}_{\text{NS-NS}} &\approx \frac{i V_{p+1}}{8\pi(8\alpha'\pi^2)^5} (4 \times 16) \int_0^\infty \frac{\pi dt}{t^2} (8\pi^2\alpha't)^{(9-p)/2} \exp\left(-\frac{ty^2}{2\pi\alpha'}\right) \\ &= i V_{p+1} 2\pi (4\pi^2\alpha')^{3-p} G_{9-p}(y), \end{aligned} \quad (12.32)$$

where $G_d(y)$ is the scalar Green's function in d -dimensions (cf. BOX 6.12)

$$G_d(y) = \frac{1}{4\pi^{d/2}} \Gamma\left(\frac{d}{2} - 1\right) y^{2-d} \equiv \int \frac{d^d k}{(2\pi)^d} \frac{e^{ik \cdot y}}{k^2}. \quad (12.33)$$

The field-theoretic computation for the graviton-dilaton potential between the p -branes is as in the bosonic case (Sect. 6.9.1) except that now the spacetime dimension D is 10 instead of 26. The amplitude is then

$$i \frac{D-2}{4} \kappa^2 \tau_p^2 G_{D-p-1}(y) = 2i \kappa^2 \tau_p^2 G_{9-p}(y), \quad (12.34)$$

where τ_p is the p -brane tension and

$$\kappa^2 = \kappa_{10}^2 e^{2\Phi}, \quad (12.35)$$

is the effective Newton constant. Comparing with (12.32), we get

$$\tau_p^2 = \frac{\pi}{\kappa^2} (4\pi^2\alpha')^{3-p}. \quad (12.36)$$

τ_p satisfies the same T -duality recursion relation as in the bosonic string, Eq. (6.295),

$$\tau_p = \tau_{p-1} (4\pi^2\alpha')^{-1/2}. \quad (12.37)$$

The RR exchange amplitude is minus the NS-NS one. The relevant terms in the low-energy action are

$$-\frac{1}{4\kappa_{10}^2} \int d^{10}x \sqrt{-G} |F^{(p+2)}|^2 + \mu_p \int_D C^{(p+1)}. \quad (12.38)$$

The kinetic term does not contain exponentials of the dilaton, so it is $1/(2\kappa_{10}^2)$ times the canonically normalized kinetic term, and the propagator of the component parallel to the D-brane is just

$$\frac{2i\kappa_{10}^2}{k^2}, \quad (12.39)$$

and the field theory amplitude becomes

$$-2i\kappa_{10}^2 \mu_p^2 G_{9-p}(y). \quad (12.40)$$

Hence,

$$\mu_p^2 = \frac{\pi}{\kappa_{10}^2} (4\pi^2 \alpha')^{3-p} = e^{2\Phi} \tau_p^2. \quad (12.41)$$

The calculation of the interactions confirms our earlier deduction that D-branes carry RR charges. We may perform a similar computation for the force between a D-brane and an orientifold plane and find that, as expected, it has an extra factor $-(2^{5-k})$: we argued above from the cancelation of divergences that the charge of the orientifold plane has a factor of $-(2^{4-k})$ (the extra factor 2 is due to the “mirror sources” produced by the orientifold).

Dirac Quantization of Charges

The D-branes RR charges satisfy a consistency relation which generalizes to higher-form gauge fields the Dirac quantization of charge in standard electromagnetism.

The original Dirac argument [12] for charge quantization was reviewed in Eqs. (8.121)–(8.124) and extended to the charges of dual electromagnetic pairs of branes of complementary dimensions. Let us specialize the argument to D-branes. A p -brane and a $(6-p)$ -brane are sources for the dual field strengths $F^{(p+2)}$ and $F^{(8-p)} \equiv *F^{(p+2)}$. We take $F^{(p+2)}$ as the basic field strength, and see the p -brane as an electric source and the $(6-p)$ -brane as a magnetic one for the gauge potential $C^{(p+1)}$. We compactify the string theory on a 6-torus $T^6 \equiv (S^1)^6$, with all circles of length 1, thus effectively making the physics four-dimensional. We take the p -brane wrapped on the first p circles of T^6 and the $(6-p)$ -brane wrapped on the remaining $(6-p)$ circles. Note that the homology classes of the two branes in $H_*(T^6)$ have intersection number 1. Moreover their volume is 1. From the viewpoint of the non-compact 4d spacetime we have just an electric and a magnetic point charge, and we can apply the Dirac argument. The conclusion was that if μ_p is the electric charge per unit volume of a p -brane and μ_{6-p} is the dual magnetic charge per unit volume carried by a $(6-p)$ -brane, one must have

$$\mu_p \mu_{6-p} = 2\pi n, \quad n \in \mathbb{Z} \quad (\text{generalized Dirac quantization}) \quad (12.42)$$

provided the normalization of charges is standard, that is, if the propagator of the gauge field is $i\eta_{\mu\nu}/k^2$. If the propagator is normalized as $i\alpha\eta_{\mu\nu}/k^2$ we get

$$\mu_p \mu_{6-p} = \frac{2\pi}{\alpha} n, \quad n \in \mathbb{Z}. \quad (12.43)$$

Note 12.4 The condition on the homology classes of the two dual branes was chosen to reflect the dual relation $F^{(8-p)} \equiv *F^{(p+2)}$ between the two sources. Indeed from the SUSY algebra (12.29) we see that for a Type II superstring compactified on a sixfold X , the effective 4d electric/magnetic charges of a Dp -brane wrapped on a p -cycle in X is given by μ_p times the homology class of the wrapped cycle. An *anti*-brane is obtained from a brane by inverting the orientation of its world-volume, and the Dirac pairing between the charges is the intersection of the associate cycles.

The Dirac relation (12.43) is satisfied by the charges (12.41) for $\alpha \equiv 2\kappa_{10}^2$. Indeed

$$\mu_p^2 \mu_{6-p}^2 = \frac{\pi^2}{\kappa_{10}^4} = \left(\frac{2\pi}{\alpha}\right)^2, \quad (12.44)$$

and the Dirac relation holds with the *minimal* quantum $n = 1$, that is,

*Type II superstring has physical objects carrying **all** electric and magnetic (RR and NS) charges allowed by Dirac quantization*

This property is known as *spectral completeness*: it is believed to be a fundamental physical principle in quantum gravity [13]. The recursion relation Eq. (6.295) and the Dirac relation (12.44) fix μ_p completely.

12.5 D-Brane Actions

We consider the world-volume effective action on a Dp -brane moving in a background of the massless bosonic fields $G_{\mu\nu}$, $B_{\mu\nu}$, Φ , $C^{(k)}$. It is a $(p+1)$ -dimensional QFT with $\mathcal{N} = 16/N(p+1)$ supersymmetry (16 supercharges) where the background fields enter as couplings for the d.o.f. propagating along the brane.

D-Branes in a NS-NS Background

The coupling of a D-brane to the NS-NS closed string fields is the same Dirac-Born-Infeld (BI) action as in the bosonic string

$$S_{D_p} = -\mu_p \int_X d^{p+1}\xi \operatorname{tr} \left\{ e^{-\Phi} \left[-\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab}) \right]^{1/2} \right\}, \quad (12.45)$$

where G_{ab} and B_{ab} are the pull-back to the world-volume X of the spacetime NS-NS fields and F_{ab} is the field strength of the gauge field living on the brane. The T -duality argument leading to this expression is the same one as for Eq. (6.274).

Recall from Sect. 6.9.1 that for n parallel D-branes with small separations, so that the strings stretched between them are light enough to be included in the low-energy action, the collective coordinates $X^m(\xi)$, gauge fields $A_a(\xi)$, and their fermionic partner $\lambda(\xi)$ get promoted to $n \times n$ matrices, i.e. the light world-volume fields transform in the adjoint of the gauge symmetry $U(n)$. In Eq. (12.45) $\text{tr}\{\cdots\}$ stands for the trace on the matrix indices. In addition there is a potential

$$V(X^m) = O([X^m, X^n]^2) \quad (12.46)$$

which is non-negative (by SUSY) and zero only for Abelian configurations. As in the bosonic case, the effect of this potential is to make the world-volume scalar fields X^m to be diagonal along its flat directions: see **BOX 12.2** for further elaboration. The v.e.v.'s of the scalars fields $X^m = \text{diag}(X_1^m, \dots, X_n^m)$ can be interpreted as the ordinary (transverse) X_i^m coordinates of n indistinguishable objects labeled by the index $i = 1, \dots, n$. For large separations $|X_i^m - X_j^m| \gg \alpha'$ ($i \neq j$) only the Abelian (diagonal) d.o.f. are relevant, since the off-diagonal ones have masses which grow linearly with the distance between the branes. At small separation these d.o.f. are light, and the full non-Abelian dynamics is relevant.

Coupling D-Branes to a RR Background

The coupling to the RR background includes terms involving the gauge field living on the brane. Like the Born–Infeld action, these terms in the action can be deduced via T -duality. Consider, for instance, a 1-brane with support on the curve $X^2 = X^2(X^1)$ in the $(1, 2)$ plane. The action contains the coupling

$$\int C^{(2)} \equiv \int dX^0 (dX^1 C_{01}^{(2)} + dX^2 C_{02}^{(2)}) = \int dX^0 dX^1 (C_{01}^{(2)} + \partial_1 X^2 C_{02}^{(2)}). \quad (12.47)$$

Under a T -duality in the 2-direction this becomes

$$\propto \int dX^0 dX^1 dX^2 (C_{012}^{(3)} + 2\pi\alpha' F_{12} C_0^{(1)}), \quad (12.48)$$

where we used the T -transformations of the RR-potentials $C^{(k)}$ as well as Eq. (6.263)

$$X'^m = -2\pi\alpha' A^m. \quad (12.49)$$

As in the bosonic string (cf. discussion around Eq. (6.277)), a D-brane forming a generic angle with the axis we T -dualize is T -dual to a D-brane with a magnetic field; in Chap. 6 this led us to the Born–Infeld action and now to the coupling of the flux F_{12} on the brane and the RR potential.

BOX 12.2 - Coulomb branch of maximal super-Yang–Mills

In the text the term “*flat directions*” stands for the submanifold C of the scalars’ field target space M of the brane world-volume theory where the (non-negative) potential $V(X)$ vanishes

$$C \stackrel{\text{def}}{=} \{X \in M : V(X) = 0\}.$$

The manifold C parametrizes the classical SUSY vacua of the world-volume theory. The classical SUSY vacua are not lifted by quantum effects (their energy remains exactly zero), so C is the manifold of *exact* quantum vacua. In a d -dimensional supersymmetric Yang–Mills theory (SYM) with 16 supercharges, as the world-volume theory of a $D(d-1)$ -brane, the vector supermultiplet contains $n = 10 - d$ real scalars X^m in the adjoint of the gauge group G . Along C we have $[X^m, X^n] = 0$, so each v.e.v. $\langle X^m \rangle$ belong to a Cartan subalgebra $\mathfrak{h}$ of the gauge group G , that is, $\langle X \rangle \in \mathfrak{h}^{10-d}$. Then, at a *generic* point of C the gauge symmetry breaks to the maximal torus

$$G \rightsquigarrow U(1)^r, \quad r = \text{rank of } G,$$

and the Yang–Mills theory is realized in its *Coulomb phase* ($\equiv$ a phase with no mass gap and massless vectors). Since the manifold C parametrizes vacua in the Coulomb phase, one refers to it as the *Coulomb branch* of SYM. Two v.e.v. $\langle X \rangle$ and $\langle X' \rangle \in \mathfrak{h}^{10-d}$ which are related by the action of the Weyl group $\text{Weyl}(G)$ are gauge equivalent, hence describe the same vacuum. Thus the Coulomb branch of the d -dimensional SYM with maximal SUSY (16 supercharges) is

$$\text{Coulomb branch of } d\text{-dimensional 16-SUSY SYM} = \mathfrak{h}^{10-d} / \text{Weyl}(G) \quad (\clubsuit)$$

where the Weyl group acts diagonally on the $(10-d)$ copies of $\mathfrak{h}$. Properly speaking, C is an *orbifold*, rather than a smooth manifold. The singular points are the non-generic loci in $X_* \in C$ where the unbroken gauge symmetry enhances with respect to the generic Abelian symmetry $U(1)^r$, to $H \times U(1)^{10-d-r_H}$; here H is the semi-simple Lie group H such that

$$\text{Weyl}(H) = \{w \in \text{Weyl}(G) : w \cdot X_* = X_*\} \equiv \begin{matrix} \text{isotropy} \\ \text{subgroup of } X_* \end{matrix}.$$

Since we have more than 8 supercharges, the geometry of the orbifold C does not receive quantum corrections [14] and its Riemannian metric (as defined by the kinetics terms in the effective Lagrangian) remains *flat*.

In the case of n parallel Dp -branes, C is the Coulomb branch of $(p+1)$ -dimensional maximal SYM with gauge group $U(n) \equiv U(1) \times SU(n)$ so Eq. $(\clubsuit)$ yields

$$C = \mathbb{R}^{n(9-p)} / \mathfrak{S}_n,$$

which parametrizes the positions of n *indistinguishable* branes in $(9-p)$ transverse directions. The Abelian subalgebra $\mathfrak{u}(1) \subset \mathfrak{u}(n)$ describes the center of mass of the brane system, a decoupled free sector from the viewpoint of the QFT on the brane. SUSY predicts that C is flat: this reflects the geometry of the spacetime where the branes are moving

Exercise 12.1 Compute the implicit overall constant in Eq. (12.48) and show that it reproduces the recursion relation on the RR charges: $\mu_p = \mu_{p-1} (4\pi^2 \alpha')^{-1/2}$.

The generalization of the last equation to an arbitrary configuration, and to multiple D-branes, produces the Chern–Simons (CS) term

$$i\mu_p \int \text{tr} \left[\exp \left(2\pi\alpha' F^{(2)} + B^{(2)} \right) \wedge \sum_q C^{(q)} \right]. \quad (12.50)$$

The expression in the integrand involves forms of different degree; the integral picks up the terms proportional to the volume form of the p -brane world-volume. There are, in addition, similar couplings where the gauge field strength $F^{(2)}$ is replaced by the spacetime curvature $R^{(2)}$: they can be computed from the superstring amplitudes.

SUSY Completion Up to now we have written only the bosonic part of the action. The fermionic terms are fixed by the 16-supercharge SUSY completion of the bosonic action.¹⁰ Writing the action as an expansion in powers of the fluctuations around a flat D-brane in flat space-time, the leading fermionic terms are given by the usual Dirac action

$$-i \int d^{p+1} \xi \text{tr} (\bar{\lambda} \Gamma^a D_a \lambda). \quad (12.51)$$

A lot is known about the full nonlinear SUSY completion of the Born–Infeld Lagrangian (the super-Born–Infeld (SBI) theory) [15]. SBI is a SUSY theory, hence has its own BPS configuration, which describe particular BPS configurations of the full string theory, e.g. fundamental strings ending on the brane [16–21]. The complete world-volume theory, which contains in addition couplings to RR fields and spacetime curvatures, together with their SUSY completions, has an even richer spectrum of BPS objects some of which will be discussed in Sects. 12.7.2 and 12.8.

Coupling Constants In perturbative field theory there is no universal definition of the coupling constants: they depend on the renormalization scheme. Even at the tree-level, there is no natural convention, and authors use definitions which differ by powers of 2 and π . In string perturbation theory the coupling is the v.e.v. of a field, $g = \exp \Phi$, but still g is not canonically defined since we are free to redefine the dilaton field by a shift $\Phi \rightsquigarrow \Phi + \text{const.}$ which multiplies g by a $O(1)$ constant. This ambiguity is unavoidable as long as g is a mere loop-counting parameter in the perturbative expansion.

However, as discussed in Sect. 8.2, in the BPS sector the dependence on the coupling is fixed by a SUSY non-renormalization theorem. This opens the possibility of giving a physical, scheme-independent, definition of the coupling which makes sense also at strong coupling, thus paving the way to non-perturbative computations in string theory. The ratio of the F-string tension to the D-string tension

$$\frac{\tau_{F1}}{\tau_{D1}} = \frac{1}{2\pi\alpha'} \frac{\kappa}{4\pi^{5/2}\alpha'} = \frac{\kappa}{8\pi^{7/2}\alpha'^2} \quad (12.52)$$

is a dimensionless physical observable proportional to the closed string coupling g . It is natural to take this ratio as the absolute definition of the string coupling

¹⁰ The existence of a SUSY completion—let alone of a maximal one with 16 supercharges—is an extremely strong condition on a bosonic action. Almost all actions do not admit SUSY completions.

$$g \stackrel{\text{def}}{=} \frac{\tau_{\text{F1}}}{\tau_{\text{D1}}} \Rightarrow \tau_{\text{D1}} = \frac{1}{2\pi g \alpha'}. \quad (12.53)$$

This definition is very convenient in the non-perturbative context. All other couplings are then expressed in terms of this ratio g and the Regge slope α' :

$$\text{Newton constant} \quad \kappa^2 = \frac{1}{2} (2\pi)^7 g^2 \alpha'^4, \quad (12.54)$$

$$\text{D-brane tension} \quad \tau_p = \frac{1}{g (2\pi)^p \alpha'^{(p+1)/2}} \quad (12.55)$$

$$\text{effective action constant} \quad \kappa_{10}^2 = \frac{1}{2} (2\pi)^7 \alpha'^4. \quad (12.56)$$

We get the coupling of the Yang–Mills theory on a D p -brane by expanding (12.45)

BOX 12.3 - Expanding the action (12.45)

One has

$$\begin{aligned} \det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab}) &= \det(G_{ab}) \cdot \det[\delta_b^a + B_b^a + 2\pi\alpha' F_b^a] = \\ &= G \cdot \exp[\text{tr} \log(1 + B + 2\pi\alpha' F)] = G \cdot \exp\left[\sum_{r=1}^{\infty} \frac{(-1)^{r-1}}{r} \text{tr}(B + 2\pi\alpha' F)^r\right] = \\ &= G \cdot \exp\left[-\frac{1}{2} \sum_{s=1}^{\infty} \frac{1}{s} \text{tr}(B + 2\pi\alpha' F)^{2s}\right] \end{aligned}$$

and

$$\begin{aligned} [-\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab})]^{1/2} &= \sqrt{-G} \cdot \exp\left[-\frac{1}{4} \sum_{s=1}^{\infty} \frac{1}{s} \text{tr}(B + 2\pi\alpha' F)^{2s}\right] = \\ &= \sqrt{-G} \left[1 - \frac{1}{4} \text{tr}(B + 2\pi\alpha' F)^2 - \right. \\ &\quad \left. - \frac{1}{32} \left(4 \text{tr}(B + 2\pi\alpha' F)^4 - (\text{tr}(B + 2\pi\alpha' F)^2)^2\right) + O((B + 2\pi\alpha' F)^6)\right] \end{aligned}$$

so the term of order F^2 in the action is

$$\frac{\mu_p}{4} \int d^{p+1}\xi \sqrt{-G} e^{-\Phi} \text{tr}(2\pi\alpha' F + B)^2,$$

and hence the effective YM coupling is

$$g_{\text{Dp}}^2 = \frac{g}{4\pi^2 \alpha'^2 \mu_p} \equiv \frac{1}{(2\pi\alpha')^2 \tau_p} = (2\pi)^{p-2} g \alpha'^{(p-3)/2}$$

$$g_{Dp}^2 = \frac{g}{4\pi^2 \alpha'^2 \mu_p} \equiv \frac{1}{(2\pi \alpha')^2 \tau_p} = (2\pi)^{p-2} g \alpha'^{(p-3)/2} \quad (12.57)$$

(see **BOX 12.3**). The coupling g_{Dp} is dimensionless for $p = 3$, and is a mass to a positive (resp. negative) power for $p < 3$ (resp. $p > 3$).

Note 12.5 The YM coupling in Type I is not equal to g_{Dp} for $p = 9$ since g_{Dp} holds in oriented theory while Type I is non-oriented.

Exercise 12.2 Show that the YM coupling in Type I is $g_{\text{YM}}^2 = 2 g_{D9}^2 \equiv 2^8 \pi^7 g \alpha'^3$.

★ D0-Brane Quantum Mechanics: New Uncertainty Relations

The D0-brane is a particle and its world-volume is a one-dimensional world-line. The effective theory living on the brane is hence a supersymmetric quantum mechanical (SQM) system which describes the particle in first quantization. At low-energy the world-line theory becomes non-relativistic, and we can use the Schrodinger picture. The non-relativistic effective Lagrangian L is the truncation of the Born–Infeld one to terms containing at most two time derivatives; L is easily obtained by T -duality starting from the 10d SYM action with the identification

$$A^i \rightarrow X^i / 2\pi \alpha'. \quad (12.58)$$

L for n D0's is just the 16-SUSY 1d SYM Lagrangian for $G = U(n)$ with a particular normalization of the 1d scalar fields X^i which identifies them with the particle position in space. The 1d scalars X^i ($i = 1, \dots, 9$) and fermions λ transform in the adjoint of the gauge group $U(n)$ and we write them as $n \times n$ Hermitian matrices. Explicitly the (non-relativistic) Lagrangian for n D0's is

$$L = \text{tr} \left\{ \frac{1}{2g\alpha'^{1/2}} D_0 X^i D_0 X^i + \frac{[X^i, X^j]^2}{4g\alpha'^{1/2}(2\pi\alpha')^2} - \frac{i}{2} \lambda D_0 \lambda + \frac{1}{4\pi\alpha'} \lambda \Gamma^0 \Gamma^i [X^i, \lambda] \right\}. \quad (12.59)$$

The first term is the non-relativistic kinetic energy with

$$m = \tau_0 \equiv 1/g\alpha'^{1/2}, \quad (12.60)$$

and we subtracted the constant rest mass $n\tau_0$. The gauge field A^0 has no kinetic term but appears in the covariant derivatives. It couples to the $U(n)$ charges, so its equation of motion enforces the constraint that physical states are $U(n)$ -invariant (*Gauss law*). The corresponding Hamiltonian is

$$H = \text{tr} \left\{ \frac{g\alpha'^{1/2}}{2} p_i p_i - \frac{1}{16\pi^2 g\alpha'^{5/2}} [X^i, X^j]^2 - \frac{1}{4\pi\alpha'} \lambda \Gamma^0 \Gamma^i [X^i, \lambda] \right\}. \quad (12.61)$$

The potential is positive since $[X^i, X^j]$ is anti-Hermitian. The canonical momentum p_i like the coordinate X^i is a $n \times n$ matrix, and the CCR read

$$[p_{iab}, X_{cd}^j] = -i\delta_i^j \delta_{ad} \delta_{bc}. \quad (12.62)$$

Now we define

$$X^i = g^{1/3} (2\pi\alpha')^{1/2} Y^i, \quad (12.63)$$

so that,

$$p_i = \frac{\Pi_i}{g^{1/3} \alpha'^{1/2}}. \quad (12.64)$$

with Π_i the conjugate momentum to Y^i . The Hamiltonian becomes

$$H = \frac{g^{1/3}}{\alpha'^{1/2}} \text{tr} \left\{ \frac{1}{2} \Pi_i \Pi_i - \frac{1}{16\pi^2} [Y^i, Y^j]^2 - \frac{1}{4\pi} \lambda \Gamma^0 \Gamma^i [Y^i, \lambda] \right\}. \quad (12.65)$$

The parameters g and α' now appear only in the overall normalization of H . It follows that the wave eigenfunctions are independent of these parameters when written in terms of the variables Y^i . The characteristic length in the wave-functions with argument the original coordinates X^i is $g^{1/3}\alpha'^{1/2}$. The length scale

$$\ell_g = g^{1/3} \sqrt{\alpha'} \quad (12.66)$$

is then a fundamental scale in the IIA theory whose physical meaning will be clear from the discussion of IIA at strong coupling Sect. 13.5. The energies scale as $g^{1/3}/\alpha'^{1/2}$ from the overall normalization of H , so the characteristic time scale as

$$\frac{1}{g^{1/3}} \sqrt{\alpha'}. \quad (12.67)$$

Putting the two together we get

$$\delta x \delta t \gtrsim \alpha' \quad (12.68)$$

which is a new stringy *uncertainty relation* involving only spacetime coordinates. In fact, one can show [22–26] that in a stringy scattering process we cannot probe spacetime positions more accurately than in (12.68). This relation gives evidence that, in string theory, the geometry of spacetime becomes non-commutative at the length scale $\sqrt{\alpha'}$, a fact that was already suggested by the fact that the brane positions become non-commuting $n \times n$ matrices X^i .

12.6 Supersymmetric Multi-brane Arrangements

A static system of D-branes, all of the same dimension p and parallel, is a half-BPS state, i.e. it preserves 16 of the 32 SUSYs. We are interested in more general BPS arrangements of D-branes which may have different dimensions or be non-parallel. They will preserve 2^k out of 32 supercharges, i.e. be 2^{k-5} -BPS. In a BPS brane arrangement there is no net force between the several branes, so the configuration remains static (i.e. the branes are relatively at rest).

12.6.1 Branes of Different Dimension Parallel to the Axes

We start by considering a system consisting of one Dp_1 -brane and one Dp_2 -brane¹¹ which are aligned to the coordinate axes, that is: each 2d field X^μ has either Neumann (N) or Dirichlet (D) b.c. at a string endpoint laying on the first (resp. second) brane. A two-brane arrangement of this form is specified by the two subsets

$$S_D^a \subset \{0, 1, \dots, 9\}, \quad a = 1, 2 \quad (12.69)$$

¹¹ We do not exclude the case $p_1 = p_2$, but now the branes may form right angles.

such that X^μ satisfies the Dirichlet b.c. at the first (resp. second) brane iff $\mu \in S_D^1$ (resp. $\mu \in S_D^2$). The complementary set of Neumann fields is written S_N^1 (resp. S_N^2). A field X^μ with $\mu \in S_D^1 \cap S_D^2$ satisfies the Dirichlet b.c. on both branes: we call it a *DD coordinate*; X^μ with $\mu \in S_N^1 \cap S_D^2$ is called a *ND coordinate*, and so on.

We ask how many supercharges are preserved by the two-brane configuration S_D^a . As always in string theory, we may study the question in different ways: from the spacetime or string world-sheet perspectives, or using the field theory living on the branes. All three approaches are useful and instructive in complementary ways.

Spacetime Viewpoint From the spacetime SUSY algebra (12.27)–(12.29) we see that the first D-brane, by itself, leaves unbroken the 16 SUSYs of the form

$$Q_\alpha + (\beta_1^\perp \tilde{Q})_\alpha, \quad \text{where } \beta_1^\perp = \prod_{m \in S_D^1} \beta^m, \quad (12.70)$$

while the second D-brane leaves unbroken the supersymmetries

$$Q_\alpha + (\beta_2^\perp \tilde{Q})_\alpha \equiv Q_\alpha + [\beta_1^\perp (\beta_1^{\perp-1} \beta_2^\perp) \tilde{Q}]_\alpha, \quad \beta_2^\perp = \prod_{m \in S_D^2} \beta^m. \quad (12.71)$$

The two-brane state is invariant only under SUSYs which are of both forms (12.70) and (12.71). These supercharges are in one-to-one correspondence with the 10d MW spinors which are left invariant by $\beta_1^{\perp-1} \beta_2^\perp$. By construction the operator $\beta_1^{\perp-1} \beta_2^\perp$ is a *reflection* in the DN and ND directions; these mixed directions form the set

$$S_D^1 \Delta S_D^1 \stackrel{\text{def}}{=} S_D^1 \cup S_D^2 \setminus S_D^1 \cap S_D^2. \quad (12.72)$$

In IIA/IIB $p_1 - p_2$ is even, and the number $\#_{ND} = \#(S_D^1 \Delta S_D^2)$ of mixed directions is also even¹² $\#_{ND} = 2j$. We may pair the mixed directions into $j \equiv \#_{ND}/2$ two-planes and write $\beta_1^{\perp-1} \beta_2^\perp$ as a product of rotations by π in these two-planes

$$\rho \equiv \beta_1^{\perp-1} \beta_2^\perp = \exp[\pi i (J_1 + J_2 + \cdots + J_j)] = \prod_{k=1}^j e^{\pi i J_k}. \quad (12.73)$$

Note that all factors in the RHS commute and may be diagonalized simultaneously. Unbroken supersymmetries correspond to eigenvectors of ρ with eigenvalue 1 acting in the $\mathbf{16}_s$ of $Spin(9, 1)$. In a spinor representation each $e^{\pi i J_k}$ has eigenvalues $\pm i$, hence: *there are unbroken supersymmetries if and only if j is even, that is, iff $\#_{ND}$ is a multiple of 4.*

When $\#_{ND} = 0 \bmod 4$ but $\#_{ND} \neq 0$, we have $\rho^2 = +1$ with $\text{tr}(\rho) = 0$, so there are 8 unbroken SUSYs: such a configuration is a $\frac{1}{4}$ -BPS state. When $\#_{ND} = 0$ (i.e. the

¹² Indeed: $\#_{ND} \equiv \#(S_D^1 \Delta S_D^2) = \#(S_D^1 \cup S_D^2) - \#(S_D^1 \cap S_D^2) = \#(S_D^1 \cup S_D^2) + \#(S_D^1 \cap S_D^2) \bmod 2 = \#S_D^1 + \#S_D^2 \bmod 2 = 18 - p_1 - p_2 \bmod 2 = p_1 - p_2 \bmod 2 = 0 \bmod 2$.

branes are parallel of the same dimension) $\rho = 1$, and we have 16 unbroken SUSYs, as in the original Type I theory which is T -dual to the $\rho = 1$ brane configuration. The number of supersymmetries should not depend on the description (the original one or its T -dual). Indeed T -duality exchanges

$$ND \leftrightarrow DN \quad \text{and} \quad NN \leftrightarrow DD \quad (12.74)$$

and leaves $\#_{ND}$ invariant.

Note 12.6 From the argument we see that the conclusion remains valid if we replace the Dp_1 by m parallel Dp_1 's and the Dp_2 by n parallel Dp_2 (all branes at rest).

World-Sheet Perspective: $p_1 - p_2$ Strings We consider the above system of Dp - Dp' branes from the world-sheet viewpoint. An open string can have both ends on the same D-brane or one end in each brane: we say that a string is of type $p_a - p_b$ if its first (second) endpoint is on the Dp_a (resp. Dp_b). The $p_1 - p_1$ and $p_2 - p_2$ string spectra were obtained before by T -duality from Type I. The $p_1 - p_2$ strings are new 2d systems that we now study.

Mode Expansions Depending on the index μ , each open string 2d field X^μ ,

$$X^\mu(w, \bar{w}) = X^\mu(w) + \tilde{X}^\mu(\bar{w}), \quad (12.75)$$

satisfies one of the four boundary conditions NN, DD, ND, or DN on the strip $0 \leq \text{Re } w \leq \pi$. As in Chap. 2, each one of the four b.c. can be enforced by the *doubling trick* (Schwarz reflection principle) in terms of either a *periodic* (**P**) or an *anti-periodic* (**A**) mode expansion

$$(\mathbf{P}) \quad X^\mu(w) = x^\mu + \left(\frac{\alpha'}{2}\right)^{1/2} \left[-\alpha_0^\mu w + \sum_{\substack{m \in \mathbb{Z} \\ m \neq 0}} \frac{\alpha_m^\mu}{m} e^{imw} \right], \quad (12.76)$$

$$(\mathbf{A}) \quad X^\mu(w) = i \left(\frac{\alpha'}{2}\right)^{1/2} \sum_{r \in \mathbb{Z} + \frac{1}{2}} \frac{\alpha_r^\mu}{r} e^{irw}, \quad (12.77)$$

by taking the *symmetric* (+) or *anti-symmetric* (−) Schwarz reflection. The anti-periodic expansion (12.77) applies to mixed coordinates ND, DN, and the periodic one (12.76) to DD, NN. When the boundary at $\text{Re } w = \pi$ has D b.c. (i.e. for DD and ND) the Schwarz reflection has an extra minus sign. The doubling tricks for the four boundary conditions are summarized in **Table 12.1**. In the R-sector the supercurrent $T_F \equiv \psi^\mu \partial X_\mu$ is periodic in the cylinder, and the Fermi field ψ^μ has the same periodicity as X^μ . In the NS sector the two periodicities are opposite.

From Eq. (12.77) we see that the 2d field X^μ in a ND/DN direction has no zero-mode. Hence p_1 - p_2 string states have *no momentum* in the DD, DN, and

Table 12.1 The Schwarz reflections for the four boundary conditions on the strip

b.c.	Mode expansion	Schwarz reflection
NN	(P) Eq. (12.76)	(+) $\tilde{X}^\mu(\tilde{w}) = X^\mu(2\pi - \tilde{w})$
DD	(P) Eq. (12.76)	(-) $\tilde{X}^\mu(\tilde{w}) = -X^\mu(2\pi - \tilde{w})$
DN	(A) Eq. (12.77)	(+) $\tilde{X}^\mu(\tilde{w}) = X^\mu(2\pi - \tilde{w})$
ND	(A) Eq. (12.77)	(-) $\tilde{X}^\mu(\tilde{w}) = -X^\mu(2\pi - \tilde{w})$

ND directions.¹³ They propagate only in the NN directions: *the p_1 - p_2 string states describe degrees of freedom which live (propagate) on the **intersection** of the two D -branes.*

Zero-Point Energy The string zero-point energy is zero in the R-sector, because the contributions from fermions and bosons cancel each other. The zero-point energy in the NS sector may be computed using the formulae in **BOX 1.2**

$$(8 - \#_{ND}) \left(-\frac{1}{24} - \frac{1}{48} \right) + \#_{ND} \left(\frac{1}{24} + \frac{1}{48} \right), \quad (12.78)$$

or by CFT techniques as

$$h_{\#_{ND}} - \frac{c}{24} \equiv \frac{\#_{ND}}{8} - \frac{1}{2}, \quad (12.79)$$

where $h_{\#_{ND}}$ is the dimension of a $Spin(2\#_{ND})$ spin-field. The oscillators raise the level by half-integers units, so only for $\#_{ND}$ a multiple of 4 NS sector masses may be equal to R-sector masses. This agrees with the spacetime analysis above: the $\#_{ND} = 2$ and $\#_{ND} = 6$ systems cannot be supersymmetric. Later we shall see that two D -branes with $\#_{ND} = 2$ do form supersymmetric *bound states*, but they form new *single-object* BPS states rather than two-brane BPS states.¹⁴

12.6.2 The World-Volume Viewpoint: The $\#_{ND} = 4$ System

The third perspective on BPS two-brane configurations is given by the world-volume theory propagating along the $\#_{NN}$ -dimensional intersection I of the two branes. From this point of view, saying that the configuration is BPS with 2^k supercharges, requires to show two things: (**1**) the light d.o.f. living on I are described by a $\mathcal{N} = 2^k / N(\#_{NN})$

¹³ A scalar field X^μ with DD boundary conditions has a zero-mode, but—as we already know from the discussion of D -branes in the bosonic string—it does not correspond to a center-of-mass momentum but rather to the separation between the two branes in that DD direction.

¹⁴ The BPS configurations in this section are BPS for all separation of the constituent branes, whereas in the $\#_{ND} = 2$ case we have bound states with non-zero binding energy.

SUSY QFT $\mathcal{T}_I$, and (2) SUSY is not spontaneously broken in the $\mathcal{T}_I$ theory. Each SUSY vacuum of $\mathcal{T}_I$ is then a 2^{k-5} -BPS extended object in spacetime.

The most important case is a Dp_1 - Dp_2 -brane pair with $\#_{\text{ND}} = 4$. There are three kinds of light strings

$$p_1 - p_1, \quad p_2 - p_2, \quad p_1 - p_2, \quad (12.80)$$

with ends in the respective branes. We focus on the case $p_1 = 5$, $p_2 = 9$ all other cases being related to this one by T -duality. We take the branes to be extended in the following directions (the symbol $\times$ means that X^μ satisfies the Neumann b.c. on the given brane)

μ	0	1	2	3	4	5	6	7	8	9
D5	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$				
D9	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

(12.81)

This configuration has a $SO(5, 1) \times SO(4)$ spacetime symmetry. The $SO(5, 1)$ group rotates the NN coordinates $\mu = 0, 1, \dots, 5$ and $SO(4)$, rotates the DN directions $\mu = 6, \dots, 9$. From the viewpoint of the theory on I , $SO(5, 1)$ is the Lorentz symmetry while $SO(4) \simeq SU(2)_R \times SU(2)'$ with $SU(2)_R$ the R-symmetry.

The 5–5 and 9–9 string states are the same one we found on an isolated D-brane. They preserve 16 supercharges and their massless bosonic sector consists of one Abelian gauge vector living on each brane and 4 scalars describing the normal motions of the D5-brane. The new d.o.f. come from the 5–9 strings: we start by studying their massless spectrum. From the spacetime and world-sheet analysis we expect the light d.o.f. to form representations of a 8-supercharge SUSY. The 5–9 d.o.f. propagate only on the intersection I of the two branes which coincides with the D5-brane world-volume since the D9 fills spacetime. The light 5–9 states describe additional massless degrees of freedom of the low-energy effective 6d QFT living on the D5 world-volume. We are interested in this effective 6d QFT which is supersymmetric with 8 supercharges.¹⁵

The NS zero-point is zero; cf. Eq. (12.79). Indeed the bosons X^μ have the expansions as in Eqs. (12.76), (12.77) (periodic for NN and DD directions, anti-periodic for the ND and DN directions). The moding of the fermions differs by $1/2$ from that of the bosons, so there are four periodic fermions ψ^m , namely those in the ND directions $m = 6, 7, 8, 9$ which transform as a vector of $SO(4)$. Their four zero-modes generate $2^{4/2} = 4$ degenerate ground states in the spin representation of $SO(4)$, which are labeled by their spins in the 6–7 and 8–9 planes

$$|s_3, s_4\rangle_{\text{NS}}, \quad s_3, s_4 = \pm \frac{1}{2}. \quad (12.82)$$

We must impose the GSO projection: taking into account the extra sign from the ghost sector, this is

¹⁵ From Eq. (8.23) we see that 6 is the largest dimension in which a field theory may have just 8 SUSYs.

$$\exp[\pi i(s_3 + s_4)] = +1 \Rightarrow s_3 = s_4. \quad (12.83)$$

We conclude that the four states (12.82) are invariant under $SO(5, 1)$ and form spinors $(\mathbf{2}, \mathbf{1}) \oplus (\mathbf{1}, \mathbf{2})$ of the internal $Spin(4) = SU(2)_R \times SU(2)'$, and only the $(\mathbf{2}, \mathbf{1})$ survives the GSO projection. Note that only the subgroup $SU(2)_R \simeq Sp(1)$ acts effectively on these massless states. This subgroup is the R-symmetry which rotates the 6d supercharges which form a symplectic Majorana–Weyl spinor (cf. Table 8.1).

In the R-sector of the *transverse* fermions ψ^i only those with $i = 2, 3, 4, 5$ are periodic, so there are again four ground states

$$|s_1, s_2\rangle_R. \quad (12.84)$$

The GSO projection does not have an extra sign¹⁶ in the R sector, so it requires

$$s_1 = -s_2. \quad (12.85)$$

The surviving spinors are invariant under the internal $SO(4)$ and form the $(\mathbf{1}, \mathbf{2})$ of the $SO(4)$ little Lorentz group of a 6d massless particle.

This system has 6d Lorentz invariance and 8 unbroken SUSYs, so we can classify its field content by 6d $(1, 0)$ supermultiplets. The massless content of the 5–9 spectrum amounts to *half* of a hypermultiplet. The other half comes from strings of the opposite orientation, of type 9–5. Neglecting higher derivative couplings, the world-volume action is fully determined by supersymmetry and the bosonic symmetries. We write its bosonic part

$$S = -\frac{1}{4g_{D9}^2} \int d^{10}x F_{MN} F^{MN} - \frac{1}{4g_{D5}^2} \int d^6x F'_{MN} F'^{MN} - \int d^6x \left[D_\mu \chi^\dagger D^\mu \chi + \frac{g_{D5}^2}{2} \sum_{A=1}^3 (\chi_i^\dagger \sigma_{ij}^A \chi_j)^2 \right]. \quad (12.86)$$

The integrals run, respectively, over the 9-brane and the 5-brane, with $M, N = 0, \dots, 9, \mu, \nu = 0, \dots, 5$, and $m = 6, \dots, 9$. The covariant derivative is

$$D_\mu = \partial_\mu + iA_\mu - iA'_\mu, \quad (12.87)$$

with A_μ and A'_μ the 9-brane and 5-brane gauge fields. The field χ_i is a $SU(2)_R$ doublet describing the hypermultiplet scalars. The 5–9 strings have one endpoint in each D-brane, so χ_i carries charges $+1$ and -1 for the respective gauge symmetries. The gauge couplings g_{Dp} were given in Eq. (12.57). In the second integral, we used the condensed notation

$$A'_M \rightarrow (A'_\mu, X'_m/2\pi\alpha'). \quad (12.88)$$

¹⁶ In our conventions.

The massless 5–5 (and also 9–9) strings separate into 6d $\mathcal{N} = 1$ vector and hypermultiplet.¹⁷ The final potential term in Eq. (12.86) is the 5–5 D-term required by supersymmetry.¹⁸ One might have expected a 9–9 D-term as well by T -duality, but this coupling is inversely proportional to the volume of the D9-brane in the (6, 7, 8, 9) directions, which has been taken to be infinite. In other words, the gauge coupling of the $U(1)$ vector associated with the D9-brane, as seen from the six-dimensional perspective, is *zero* and the corresponding $U(1)$ symmetry is realized as a *flavor* symmetry of the 6d effective field theory.

Under T -dualities in the ND directions, one obtains

$$(p_1, p_2) = (8, 6), (7, 7), (6, 8), (5, 9), \quad (12.89)$$

but the intersection I of the two branes remains $(5 + 1)$ -dimensional and we still have the action (12.86) on I . T -dualities in the NN directions give

$$(p_1, p_2) = (9 - r, 5 - r). \quad (12.90)$$

The vector components in the T -dualized directions become collective coordinates as usual

$$A_i \rightarrow X_i/2\pi\alpha', \quad A'_i \rightarrow X'_i/2\pi\alpha'. \quad (12.91)$$

The term $D_i \chi^\dagger D^i \chi$ then becomes

$$\left(\frac{X_i - X_j}{2\pi\alpha'} \right)^2 \chi^\dagger \chi. \quad (12.92)$$

This just reflects the fact that the $(9 - r)$ brane and the $(5 - r)$ brane may be separated in the DD directions and then the strings stretched between them become massive.

Non-Abelian Generalization The low-energy effective action for several branes of each type is given by the obvious non-Abelian extension of the above ones. If we have m D5 and n D9 the resulting hypermultiplet will transform in the representation $\mathbf{m}$ of a gauge $U(m)$ group and in the representation $\mathbf{n}$ of a flavour $U(n)$.

¹⁷ For a single brane they are the vector-multiplet of 6d (1, 1) SUSY which decomposes under 6d (1, 0) SUSY into one vector-multiplet and one hypermultiplet.

¹⁸ The theory has 8 supercharges, and so it behaves like $\mathcal{N} = 2$ in 4d. The $U(1)$ vector multiplets as a real $SU(2)_R$ triplet of auxiliary fields Y_{ij} which decomposes into the auxiliary field D , F , F^* in terms of the $\mathcal{N} = 1$ vector+chiral supermultiplets. Integrating out Y_{ij} produces this quartic term in the superpotential.

12.6.3 Non-parallel Branes

Next we study the case of Dp -branes rotated at general angles. We start with a static configuration of k parallel Dp -branes which are separated in the direction X^9 and extended along the vector subspace $\mathbb{R}^p \equiv \{X^{p+1} = \dots = X^8 = 0\} \subset \mathbb{R}^8$ of the transverse $\mathbb{R}^8$. This configuration is invariant under 16 SUSYs. Keeping fixed the first brane, we rotate the i th brane by $\rho_i \in SO(8)$ in the transverse $\mathbb{R}^8$. How many SUSY are preserved by the brane configuration for given rotations $\{\rho_i\}$?

We write ϱ_i for ρ_i seen as a 16×16 matrix in the spinorial representation $\mathbf{16}_{\text{spin}} \equiv \mathbf{8}_s \oplus \mathbf{8}_c$. Rotating the N or D directions between themselves does not change the brane configuration: the resulting brane arrangement depends only on the class of ϱ_i in the coset $Spin(8)/[Spin(p) \times Spin(8-p)]$. $Spin(p) \times Spin(8-p)$ is the subgroup of $Spin(8)$ which commutes with $\beta^\perp \equiv \prod_{m=p+1}^9 \beta^m$, and we are free to choose the coset representatives ϱ_i such that

$$\varrho_i \beta^\perp = \beta^\perp \varrho_i^{-1} \quad (12.93)$$

as the parametrization of the final configuration.

Spacetime Viewpoint

The spacetime perspective leads to a simple and elegant answer:

Fact 12.1 Let $H \subset Spin(8)$ be the smallest Lie subgroup of the form in **Corollary 11.3** (i.e. a holonomy group consistent with parallel spinors in dimension 8) which contains all the¹⁹ ϱ_i^2 's (with $\varrho_1^2 = 1$). If no such group contains all ϱ_i^2 's we set $H = Spin(8)$. The number of SUSYs preserved by the brane configuration is equal to the number of parallel spinors for H as given by Eq. (11.19) with $(d+n)$ equal 8 minus the dimension of the manifold M of holonomy H (first column in the table of **Theorem 11.5**).

The inclusions between the groups in **Fact 12.1** are

$$\begin{array}{ccccccc}
 & & & G_2 & \longrightarrow & Spin(7) & \\
 & & & \uparrow & & \uparrow & \searrow \\
 1 & \longrightarrow & SU(2) & \longrightarrow & SU(3) & \longrightarrow & SU(4) \longrightarrow Spin(8) \\
 & & \searrow & & & \nearrow & \\
 & & & SU(2) \times SU(2) & & &
 \end{array} \quad (12.94)$$

and we have one supercharge for $H = Spin(7)$, G_2 , and $2^{4-\sum_i (n_i-1)}$ supercharges for $H = \prod_i SU(n_i)$. When $H = Spin(8)$ all SUSYs are broken.

We present several proofs of **Claim 12.1** from different perspectives which teach us important lessons on the brane arrangement. We begin with the spacetime SUSY

¹⁹ A priori ϱ_i is determined only up to sign. However only ϱ_i^2 enters the argument, and the sign ambiguity cancels out.

algebra viewpoint. The first brane is invariant under the supersymmetries of the form $Q_\alpha + (\beta^\perp \tilde{Q})_\alpha$, while the i th brane by $Q_\alpha + (\varrho_i^{-1} \beta^\perp \varrho_i \tilde{Q})_\alpha$. The $\mathbf{16}_s$ of $SO(9, 1)$ decomposes into $\mathbf{8}_s \oplus \mathbf{8}_c$ of the subgroup $SO(8)$, and the unbroken SUSY correspond to the elements of $\mathbf{8}_s \oplus \mathbf{8}_c$ which are left invariant by the elements of $Spin(8)$

$$(\beta^\perp)^{-1} \varrho_i^{-1} \beta^\perp \varrho_i \equiv \varrho_i^2, \quad \varrho_i^{-1} (\beta^\perp)^{-1} \varrho_i \varrho_j^{-1} \beta^\perp \varrho_j = \varrho_i^{-2} \varrho_j^2, \quad (12.95)$$

i.e. by the group generated by the ϱ_i^2 . If this group is contained in a subgroup $H \subset Spin(8)$ leaving invariant 2^k spinors, we have 2^k unbroken supercharges, i.e. the brane arrangement is 2^{k-5} -BPS. The list of the relevant subgroups is as in Eq. (12.94). We illustrate the result in a simple configuration.

Example: Two Rotated D4-Branes

Following refs. [27, 28] we consider two D4-branes initially extended in the (2,4,6,8)-directions, and separated by some distance y_1 in the 1-direction. We rotate the second brane by an angle ϕ_a in the $2a-2a+1$ plane for $a = 1, \dots, 4$. The resulting rotation $\varrho^2 \in Spin(8)$ is

$$\varrho^2 \equiv \exp \left(2i \sum_{a=1}^4 s_a \phi_a \right) \quad s_a = \pm \frac{1}{2}. \quad (12.96)$$

In the $\mathbf{16}_s$ of $SO(9, 1)$ the numbers $(2s_1, 2s_2, 2s_3, 2s_4)$ run over all combinations of ± 1 . Each combination such that the phase (12.96) is 1 gives an unbroken SUSY. For generic $\phi_a, \varrho \in U(4)$ and there is no unbroken SUSY. For angles $\phi_1 + \phi_2 + \phi_3 + \phi_4 = 0 \pmod{2\pi}$ (but otherwise generic) there are two unbroken SUSY, the ones with $s_1 = s_2 = s_3 = s_4$: indeed in this case $\varrho \in SU(4)$. There are 4 SUSY when $\phi_1 + \phi_2 + \phi_3 = \phi_4 = 0 \pmod{2\pi}$ or when $\phi_1 + \phi_2 = \phi_3 + \phi_4 = 0 \pmod{2\pi}$ which correspond, respectively, to $\varrho \in SU(3)$ and $\varrho \in SU(2) \times SU(2)$. For $\phi_1 + \phi_2 = \phi_3 = \phi_4 = 0 \pmod{2\pi}$ there are 8 unbroken SUSY, and $\varrho \in SU(2)$.

Fact 12.1: String World-Sheet Viewpoint

For simplicity we focus on the set-up in the **Example**: two D4-branes separated along the first axis and rotated in the 4 two-planes 2–3, 4–5, 6–7, and 8–9 by, respectively, angles $\phi_1, \phi_2, \phi_3, \phi_4$. We introduce complex coordinates in the 4 two-planes

$$Z^a = X^{2a} + i X^{2a+1}, \quad a = 1, \dots, 4 \quad (12.97)$$

so that ρ is the rotation

$$\rho: Z^a \mapsto e^{i\phi_a} Z^a. \quad (12.98)$$

! Warning

Equation (12.97) is a choice of complex structure in $\mathbb{R}^8$. Flipping a coordinate sign $X^{2a+1} \leftrightarrow -X^{2a+1}$ gives an inequivalent choice with $\phi_a \leftrightarrow -\phi_a$. Spacetime parity is a symmetry of IIA, so the physics cannot depend on the choice.

The transverse $SO(8)$ rotation group contains a $U(4)$ subgroup which preserves the complex structure (12.97). The rotation ρ is the $U(4)$ matrix

$$\rho = \text{diag}(e^{i\phi_a}). \quad (12.99)$$

Fact 12.1 states that, when $\rho \in \prod_k SU(n_k) \subset U(4)$ the two-brane system is invariant under $2^{4-\sum_k(n_k-1)}$ SUSYs (cf. Eq. (12.96)). We wish to recover this result from the world-sheet viewpoint. For clarity we start with a Minkowskian world-sheet; at the end we shall Wick rotate back to Euclidean signature. We consider strings suspended between the two branes: the endpoint $\sigma^1 = 0$ is on the unrotated brane and the endpoint $\sigma^1 = \pi$ on the rotated one, where (σ_1, σ_2) are the space and time coordinates in the Minkowskian strip. The b.c. for the complex scalar field Z^a are

$$\sigma^1 = 0 \quad \partial_1 \text{Re}(Z^a) = \text{Im}(Z^a) = 0 \quad (12.100)$$

$$\sigma^1 = \pi \quad \partial_1 \text{Re}(e^{-i\phi_a} Z^a) = \text{Im}(e^{-i\phi_a} Z^a) = 0, \quad (12.101)$$

which are automatically satisfied by

$$Z^a(\sigma_1, \sigma_2) = \mathcal{Z}^a(\sigma_1 + \sigma_2) + \overline{\mathcal{Z}^a}(-\sigma_1 + \sigma_2), \quad (12.102)$$

where $\mathcal{Z}^a: \mathbb{R} \rightarrow \mathbb{C}$ is an analytic function satisfying the periodicity condition

$$\mathcal{Z}^a(w + 2\pi) = e^{2i\phi_a} \mathcal{Z}^a(w), \quad (12.103)$$

and the bar is complex conjugation. We set $\theta_a = \phi_a/\pi$; we stress that Eq. (12.103) depends on θ_a only mod 1. The mode expansions have the form

$$\begin{aligned} \mathcal{Z}^a(\sigma_1 + \sigma_2) &= \sum_{r \in \mathbb{Z}} \alpha_r^a e^{i(r+\theta_a)(\sigma_1+\sigma_2)}, \\ \overline{\mathcal{Z}^a}(-\sigma_1 + \sigma_2) &= \sum_{r \in \mathbb{Z}} \bar{\alpha}_r^a e^{i(r+\theta_a)(\sigma_1-\sigma_2)}. \end{aligned} \quad (12.104)$$

To get the Euclidean formulation, we regard $\mathcal{Z}^a(w)$ as a holomorphic function of its complex argument w which satisfies the periodicity condition (12.103). Using the formulae in **BOX 1.2**, we see that the “vacuum energy” in the R-sector is zero, since the bosonic and fermionic fields have the same b.c. The GSO-allowed R ground states then correspond to fermions in spacetime with mass

$$m^2 = \frac{y_1^2}{(2\pi\alpha')^2} \quad (12.105)$$

from the tension of the stretched string. If SUSY is unbroken, we must have in the spectrum bosons with the same mass. For the NS sector, the four complex bosons contribute to the shift constant $a_{\text{NS}} = a_{\text{bos}} + a_{\text{fer}}$ a quantity

$$a_{\text{bos}} = -\frac{4}{12} + \frac{1}{2} \sum_a \theta_a (1 - \theta_a). \quad (12.106)$$

a_{bos} is invariant as $\theta_a \leftrightarrow 1 - \theta_a$ which corresponds to $Z^a \leftrightarrow \bar{Z}^a$ (a different choice of complex structure) followed by an integral shift of θ_a . Note that a_{bos} is *not* periodic in θ_a , that is, a_{bos} is a *multivalued* function of the phase $\exp(2\pi i \theta_a) \in S^1$ in (12.103) due to the *spectral flow* phenomenon. The contribution from the fermions is

$$a_{\text{fer}} = -\frac{4}{24} + \frac{1}{2} \sum_a \theta_a^2 \quad (12.107)$$

which is invariant under the flip $\theta_a \leftrightarrow -\theta_a$ *without* shifts. The function $a_{\text{fer}}(\theta_a)$ also has multiple determinations because of the spectral flow. The function $a_{\text{NS}}(\phi_a)$ then has several branches, and one should be careful to select the physical sheet appropriate for each brane configuration. In the first branch the NS shift constant is

$$a_{\text{NS}} \equiv a_{\text{bos}} + a_{\text{fer}} = -\frac{1}{2} + \frac{1}{2} \sum_a \theta_a \quad (12.108)$$

while the moding of the fermions $\psi_{n+1/2+\theta_a}^a$ and $\bar{\psi}_{n+1/2-\theta_a}^a$ with $n \in \mathbb{Z}$. Hence the GSO-allowed bosonic states

$$\psi_{-\frac{1}{2}+\theta_a}^a |0; k\rangle_{\text{NS}} \quad \text{have} \quad m^2 = \frac{y_1^2}{(2\pi\alpha')^2} + \frac{1}{2\pi\alpha'} \left(-2\phi_a + \sum_b \phi_b \right) \quad (12.109)$$

and we have spacetime bosons degenerate in mass with the R ground states iff the phases (12.96), with 3 s_a of the same sign are equal to 1. This shows that the spectrum is supersymmetric when 8 out of the 16 phases $\exp(\pm i\phi_1 \pm i\phi_2 \pm i\phi_3 \pm i\phi_4)$ in Eq. (12.96) are equal 1. The remaining 8 phases are obtained by going to the appropriate branch of the function $a_{\text{NS}}(\phi_a)$ obtained by flipping the sign of one of the angles $\phi_a \leftrightarrow -\phi_a$ which in particular entails $\psi_{n+1/2+\theta_a}^a \leftrightarrow \bar{\psi}_{n+1/2-\theta_a}^a$.

Tachyons in Brane Configurations In a non-SUSY brane configuration a tachyon can appear. In Eq. (12.109) with $a = 1$ set $\phi_2 = \phi_3 = \phi_4$ and $0 < \phi_1 \leq \pi$. The state

$$\psi_{-\frac{1}{2}+\theta_1}^1 |0; k\rangle_{\text{NS}} \quad \text{has mass}^2 \quad m^2 = \frac{y_1^2}{4\pi^2\alpha'^2} - \frac{\phi_1}{2\pi\alpha'}. \quad (12.110)$$

This is negative for separations y_1 below the critical value $y_1^c = \sqrt{2\pi\alpha'\phi_1}$. The physical interpretation of this result is clear: consider for simplicity the extreme case when $\phi_1 = \pi$, i.e. when the branes are anti-parallel rather than parallel, both the NS-NS and R-R exchanges are attractive,²⁰ and the configuration is unstable against

²⁰ The RR charge is a p -form, so it flips sign under inversion of the orientation: parallel branes have the same charge (they repulse each other), anti-parallel branes have opposite charge (they attract).

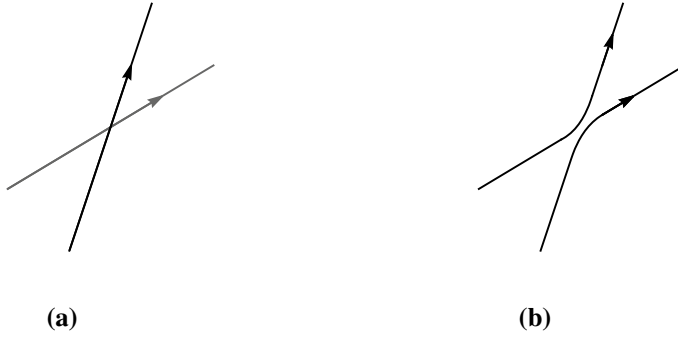


Fig. 12.1 **a** Two D4-branes forming an angle $\theta < \pi$. If their separation is smaller than a certain critical length, the two-brane system has tachyonic modes, so is unstable and evolves toward a lower energy configuration. **b** The final state of lower energy of the two-brane system in Fig. 12.1a, obtained by a breaking/gluing recombination of the branes. At infinity the configuration is identical to the one in Fig. 12.1a

the process of brane/anti-brane annihilation. Even when the D-branes are almost (but not exactly) parallel and close enough, as in Fig. 12.1a, they may lower their energy by recombining themselves as in Fig. 12.1b. The recombination process is the origin of the instability. In this situation the tachyon has a natural physical interpretation: it does not signal an inconsistency of the theory, but rather the (expected!) fact that non-SUSY configurations tend to be unstable.

Force Between Rotated D-Branes

We continue with the set-up in the **Example**. For generic angles ϕ_a the two D4-brane configuration is non-BPS and there is a net force between them that we wish to compute from the string amplitude on the cylinder with one boundary on each of the two branes (cf. Fig. 6.3). The cylinder amplitude involves traces of $q^{L_0 - c/24}$ over the states of the string suspended between the two rotated branes. The world-sheet theory satisfies the boundary conditions (12.100), (12.101) with mode expansions the Wick rotation of (12.104). The partition function for one such complex scalar is

$$q^{E_0(\phi)} \prod_{m=0}^{\infty} [1 - q^{m+\phi/\pi}]^{-1} [1 - q^{m+1-\phi/\pi}]^{-1} = i \frac{e^{\phi^2 t/\pi} \eta(it)}{\theta_{11}(i\phi t/\pi, it)} \quad (12.111)$$

where $q = \exp(-2\pi t)$ and

$$E_0(\phi) = -\frac{1}{12} + \frac{1}{2} \frac{\phi}{\pi} \left(1 - \frac{\phi}{\pi}\right) \equiv \frac{1}{24} - \frac{1}{2} \left(\frac{\phi}{\pi} - \frac{1}{2}\right)^2 \quad (12.112)$$

BOX 12.4 - Some useful formulae for θ -functions

As before, for the half-integral characteristics we write

$$\theta_{a,b}(y, \tau) \equiv \vartheta \left[\begin{matrix} a/2 \\ b/2 \end{matrix} \right] (y, \tau) \quad a, b = 0, 1.$$

From **BOX 5.5** we have the Jacobi triple product identity

$$\vartheta \left[\begin{matrix} a \\ b \end{matrix} \right] (y, \tau) = \eta(\tau) e^{2\pi i a(b+y)} q^{a^2/2-1/24} \prod_{n=1}^{\infty} (1 + q^{n+a-1/2} e^{2\pi i(b+y)}) (1 + q^{n-a-1/2} e^{-2\pi i(b+y)}),$$

which gives (for $z = e^{2\pi i y}$)

$$\theta_{1,1}(y, \tau) = -2 \sin(\pi y) \eta(\tau) q^{1/12} \prod_{n=1}^{\infty} (1 - z q^n) (1 - z^{-1} q^n). \quad (\clubsuit)$$

In particular,

$$\theta_{1,1}(y/\tau, -1/\tau) = -i \sqrt{-i\tau} \exp(i\pi y^2/\tau) \theta_{1,1}(y, \tau). \quad (\diamond)$$

is the zero-point energy; cf. **BOX 1.2** or Eq. (12.106). Equation (12.111) follows from the Jacobi triple product representation of the θ -functions, see **BOX 12.4**.

Similarly, in each of the sectors of the fermionic path integral one replaces the function $Z_{\beta}^{\alpha}(it)$ which appears for parallel D-branes with its generalization defined and studied in Sect. 6.3 (cf. Eq. (6.73))

$$Z_{\beta}^{\alpha}(\phi, it) = \frac{\theta_{\alpha,\beta}(i\phi t/\pi, it)}{\exp(\phi^2 t/\pi) \eta(it)}. \quad (12.113)$$

For $\alpha = \beta = 1$ this is the inverse²¹ of (12.111) since the partition function for complex fermions is a determinant while for complex scalars is an inverse determinant. There is again a subtlety with the choice of branch for the zero-energy as a multivalued function of the angle ϕ . The full fermionic partition function is then

$$\frac{1}{2} \left[\prod_{a=1}^4 Z_0^0(\phi_a, it) - \prod_{a=1}^4 Z_0^1(\phi_a, it) - \prod_{a=1}^4 Z_1^0(\phi_a, it) \mp \prod_{a=1}^4 Z_1^1(\phi_a, it) \right] \quad (12.114)$$

which generalizes $Z_{\psi}^{\pm}(it)$. The two signs are equivalent, being related by a parity transformation in target space $\phi_4 \leftrightarrow -\phi_4$. Although our usual convention is to pick the upper sign, we get more canonically looking formulae by choosing the lower sign. Then Eq. (12.114) becomes the RHS of the Riemann relation (6.93) for $\alpha = \beta = 1$ and $y_a = i\phi_a t/\pi$. The (transverse) fermionic partition function reduces to

²¹ Up to a fourth root of unity which cancels out in the final expression (12.114).

$$\prod_{a=1}^4 Z_1^1(\phi'_a, it) \quad \text{where} \quad \phi'_a \stackrel{\text{def}}{=} S_{ab}\phi_b. \quad (12.115)$$

By definition of the matrix S (cf. Sect. 6.3), the ϕ'_a are the eigenvalues of the $SO(8)$ rotations of angles ϕ_a in the representation (s) of $Spin(8)$ since the linear map $S: \phi_a \mapsto \phi'_a$ is the triality automorphism which sends vector weights to (s) weights.

Collecting all factors, the potential is

$$V = - \int_0^\infty \frac{dt}{t} (8\pi^2 \alpha' t)^{-1/2} \exp\left(-\frac{ty_1^2}{2\pi\alpha'}\right) \prod_{a=1}^4 \frac{\theta_{1,1}(i\phi'_a t/\pi, it)}{\theta_{1,1}(i\phi_a t/\pi, it)}. \quad (12.116)$$

The function $\theta_{1,1}(z, it)$ is odd in z and vanishes at $z = 0$. If any of the ϕ_a vanish, the denominator is zero and the amplitude diverges. This is due to a bosonic zero-mode whose physical meaning is obvious: when $\phi_a = 0$ the branes are parallel in one direction and translation in that direction is a zero-mode. We must replace

$$i\theta_{1,1}(i\phi_a t/\pi, it)^{-1} \rightarrow \eta(it)^{-3} (8\pi^2 \alpha' t)^{-1/2}, \quad (12.117)$$

which is the proper factor *per unit length* for a non-compact direction.

Exercise 12.3 Take $\phi_4 \rightarrow 0$ so the D4-branes extend in the 8-direction, and T -dualize in this direction. Write the potential for the resulting pair of rotated D3-branes.

When some ϕ'_a vanish, the potential V is zero. The reason is that there is an unbroken SUSY: the 8 phases $\exp(\pm 2i\phi'_a)$ are a subset of the 16 phases (12.96) at which the two D4 configuration is BPS. Again, the potential V is a multi-branch function of the angles, and the other 8 BPS angles are recovered in a different sheet of the potential obtained by target-space parity. One should be careful which determination of V is the physically correct one for each brane configuration.

Long Distance Asymptotics We consider the long distance limit, $y_1 \rightarrow \infty$, of V . The exponential factor in (12.116) forces t to be small, and the θ -functions become

$$\prod_{a=1}^4 \frac{\theta_{1,1}(i\phi'_a t/\pi, it)}{\theta_{1,1}(i\phi_a t/\pi, it)} \rightarrow \prod_{a=1}^4 \frac{\sin \phi'_a}{\sin \phi_a} \quad (12.118)$$

see **BOX 12.5**. More generally, one has an expansion

$$\prod_{a=1}^4 \frac{\theta_{1,1}(i\phi'_a t/\pi, it)}{\theta_{1,1}(i\phi_a t/\pi, it)} = \left(\prod_{a=1}^4 \frac{\sin \phi'_a}{\sin \phi_a} \right) \left(1 + \sum_{n \geq 1} f_n(\phi_a, \phi'_a) e^{-2\pi n/t} \right), \quad (12.119)$$

BOX 12.5 - Proof of (12.118)

Setting $\tau = i/t$ in Eq. ($\diamond$) of the box on page 684,

$$\theta_{1,1}(-iyt, it) = -i t^{-1/2} \exp(\pi y^2 t) \theta_{1,1}(y, i/t),$$

while

$$\theta_{1,1}(y, i/t) \Big|_{t \rightarrow 0} \sim -2 \sin(\pi y) e^{-\pi/(4t)} \left(1 + O(e^{-2\pi/t})\right). \quad (\spadesuit\spadesuit)$$

Hence

$$\frac{\theta_{1,1}(iyt/\pi, it)}{\theta_{1,1}(izt/\pi, it)} \Big|_{t \rightarrow 0} = \frac{\sin y}{\sin z}.$$

for certain functions f_n of the angles. The t -integral then gives a power of the separation for the leading term. The result agrees with the low-energy field theory calculation. For 4-branes with all ϕ_a non-zero we get a potential linear in y_1

$$V = \frac{y_1}{2\pi\alpha'} \prod_{a=1}^4 \frac{\sin \phi'_a}{\sin \phi_a} + O(e^{-2|y_1|/\sqrt{\alpha'}}). \quad (12.120)$$

Applying T -duality, we get the force between rotated Dp-branes for $p \neq 4$. For instance we set $\phi_4 = 0$, so that both 4-branes are parallel to the 8-axis, and T -dualize in this direction to get two D3-branes rotated in the first three 2-planes. The fermionic partition function is unaffected, while the factors (12.117) are replaced by

$$\eta(it)^{-3} \exp\left[-\frac{t(y_8^2 + y_9^2)}{2\pi\alpha'}\right]. \quad (12.121)$$

For generic angles ϕ_a the potential between D3-branes falls as $1/y_1$: indeed from Eq. (12.121) one replaces the LHS of (12.118) with

$$-i \left(\prod_{a=1}^3 \frac{\theta_{1,1}(i\phi'_a t/\pi, it)}{\theta_{1,1}(i\phi_a t/\pi, it)} \right) \frac{\theta_{1,1}(i\phi'_4 t/\pi, it)}{\eta(it)^3} \quad (12.122)$$

since

$$\eta(it) \simeq t^{-1/2} e^{-\pi/(12t)} \left(1 + O(e^{-2\pi/t})\right) \quad \text{as } t \rightarrow 0, \quad (12.123)$$

we have

$$-i \left(\prod_{a=1}^3 \frac{\theta_{1,1}(i\phi'_a t/\pi, it)}{\theta_{1,1}(i\phi_a t/\pi, it)} \right) \frac{\theta_{1,1}(i\phi'_4 t/\pi, it)}{\eta(it)^3} \Big|_{t \sim 0} \simeq 2t \sin \phi'_4 \prod_{a=1}^3 \frac{\sin \phi'_a}{\sin \phi_a} \quad (12.124)$$

and

$$V = -\frac{1}{y} \sin \phi'_4 \prod_{a=1}^3 \frac{\sin \phi'_a}{\sin \phi_a} + \text{exponentially small}, \quad (12.125)$$

where $y^2 = y_1^2 + y_8^2 + y_9^2$. In the same fashion, for all of rotated p -branes we get the power-law expected from field theory, i.e. V is proportional to the Coulomb potential in a space of dimension equal to the number of transverse directions

$$V \propto \frac{1}{y^{7-2p}}. \quad (12.126)$$

Exercise 12.4 Prove the last assertion.

★ D-Brane Scattering

Static parallel D-branes are BPS and hence there is no net force between them. When they are in relative motion all SUSY is broken²² and there is a velocity-dependent force between the branes. Since the Lorentz boost is the Wick rotation of a spatial rotation, the velocity-dependent force is simply the analytic continuation of the static potential for rotated branes computed above. To compute the force between parallel branes in relative motion, we consider the case in which $\phi_a = 0$ for $a \geq 2$. The rotated brane then satisfies

$$X^3 = X^2 \tan \phi_1. \quad (12.127)$$

We analytically continue $X^2 \rightarrow iX^{0'}$ and $\phi_1 \rightarrow -iu$. Then

$$X^3 = X^{0'} \tanh u, \quad (12.128)$$

which describes a D4-brane moving in the 3-direction with a constant velocity $v = \tanh u$ in the time $X^{0'}$. Next we Wick rotate $X^0 \rightarrow -iX^{2'}$ to eliminate the extra time coordinate.

The interaction (12.116) between the D-branes becomes

$$\mathcal{A} = -iV_p \int_0^\infty \frac{dt}{t} (8\pi^2 \alpha' t)^{-p/2} \exp\left(-\frac{ty_1^2}{2\pi\alpha'}\right) \frac{\theta_{1,1}(ut/2\pi, it)^4}{\eta(it)^9 \theta_{1,1}(ut/\pi, it)}, \quad (12.129)$$

where we have written the result for general p by taking T -duality along the appropriate axes.

We rewrite the expression in a more illuminating form by a modular transformation

$$\mathcal{A} = \frac{V_p}{(8\pi^2 \alpha')^{p/2}} \int_0^\infty \frac{dt}{t} t^{(6-p)/2} \exp\left(-\frac{ty_1^2}{2\pi\alpha'}\right) \frac{\theta_{1,1}(iu/2\pi, i/t)^4}{\eta(it)^9 \theta_{1,1}(iu/\pi, i/t)}. \quad (12.130)$$

We can rewrite this amplitude as an integral over the effective world-line

$$\mathcal{A} = -i \int_{-\infty}^{+\infty} d\tau V(r(\tau), v) \quad (12.131)$$

where

$$r(\tau)^2 = y_1^2 + v^2 \tau^2, \quad v = \tanh u, \quad (12.132)$$

²² The analysis is the Wick rotation of the one in Sect. 12.6.1. Now $\rho \in \text{Spin}(9, 1)$ is a Lorentz boost rather than a $\text{Spin}(8)$ rotation, i.e. and element of the subgroup $\text{Spin}(1, 1) \subset \text{Spin}(9, 1)$, isomorphic to $\mathbb{R}$. The eigenvalues of ρ^2 are $e^{\pm\eta}$ (where $\eta \in \mathbb{R}$ is the rapidity of the boost) which may be 1 only for zero rapidity.

and

$$V(r, v) = i \frac{2V_p}{(8\pi^2\alpha')^{(p+1)/2}} \int_0^\infty dt t^{(5-p)/2} \exp\left(-\frac{tr^2}{2\pi\alpha'}\right) \frac{(\tanh u) \theta_{1,1}(iu/2\pi, i/t)^4}{\eta(it)^9 \theta_{1,1}(iu/\pi, i/t)}. \quad (12.133)$$

To get (12.131), note that the only factor in the integrand of $V(r, v)$ which contains r , hence τ , is the exponential one. The integral in τ reduces to the Gaussian integral

$$\int_{-\infty}^{+\infty} d\tau \exp\left(-\frac{ty_1^2 + tv^2\tau^2}{2\pi\alpha'}\right) = \frac{\sqrt{8\pi^2\alpha'}}{2v\sqrt{t}} e^{-ty_1^2/2\pi\alpha'}. \quad (12.134)$$

It is instructive to consider the behavior of $V(r, v)$ as $v \rightarrow 0$ (i.e. $u \rightarrow 0$) $V(r, v)$ vanishes as v^4 because of the zeros of the θ -functions. Time-reversal symmetry predicts that only even powers of v are present. The vanishing of $O(v^0)$ and $O(v^2)$ terms follows from the SUSY non-renormalization theorems of Chap. 10. The low-energy field theory of the D-branes is a $U(1) \times U(1)$ supersymmetric gauge theory with 16 supercharges. What we are computing is a correction to the world-volume effective action arising from integrating out massive states, i.e. strings stretched between the D-branes. The vanishing of the v^2 term follows from the result in Chap. 10 that for all theories with more than 8 supercharges all corrections to interactions with at most 2 derivatives are forbidden. Instead, if we take $\phi_3 = \phi_4 = \pi/2$, so that $\#_{\text{NS}} = 4$, we have an effective theory with only 8 supercharges and $O(v^2)$ corrections are allowed: indeed in this case the numerator has a only a double zero, so the potential is $O(v^2)$.

The interaction $V(r, v)$ is in a complicated function of the separation y_1 , but in an expansion in powers of the velocity the leading $O(v^4)$ term is simple; from, say, Eq. (♣) in BOX 12.4, we have

$$\theta_{1,1}(y, \tau) = -2\pi y \eta(\tau)^3 + O(y^2), \quad (12.135)$$

that is,

$$2i \frac{(\tanh u) \theta_{1,1}(iu/2\pi, i/t)^4}{\eta(it)^9 \theta_{1,1}(iu/\pi, i/t)} = -u^4 + O(u^4) \equiv -v^4 + O(v^6). \quad (12.136)$$

Then

$$\begin{aligned} V(r, v) &= -v^4 \frac{V_p}{(8\pi^2\alpha')^{(p+1)/2}} \int_0^\infty dt t^{(5-p)/2} \exp\left(-\frac{tr^2}{2\pi\alpha'}\right) + O(v^6) = \\ &= -\frac{v^4}{r^{7-p}} \frac{V_p}{(\alpha')^{p-3}} 2^{1-2p} \pi^{(5-3p)/2} \Gamma\left(\frac{7-p}{2}\right) + O(v^6). \end{aligned} \quad (12.137)$$

At long distance this result is in agreement with supergravity.

12.7 BPS Bound States of Branes

The BPS objects yield the main clue to understand the physics of supersymmetric string theories beyond weak coupling. The detailed study of BPS bound states of D-branes, and D-branes with fundamental strings (F-strings), is a crucial step toward understanding superstring theory at strong coupling [29, 30].

The branes are extended objects. To make everything well-defined, we need to make their volume finite: we always think the p -branes as wrapped around a *very large* p -torus T^p , i.e. the i th Neumann direction tangent to be brane is taken to be

periodic of period $L_i \gg \sqrt{\alpha'}$. If we have several branes of different dimensions $p' \leq p$, each p' -brane is wrapped on a p' -cycle in T^p . From the perspective of the non-compact $10 - p$ dimensions, a p' -brane then looks like a particle of mass

$$m = (\text{tension of } p'\text{-brane}) \times (\text{volume of wrapped } p'\text{-cycle}), \quad (12.138)$$

with a RR charge proportional to the homology class of the wrapped p' -cycle. A state in the system of several branes is then a *bound state* iff it is a *normalizable eigenstate* of m^2 . The BPS configurations of several parallel branes with $\#_{\text{ND}} = 0 \bmod 4$ are not normalizable eigenstates of m^2 since there is no potential for their relative separation in the non-compact directions x^μ so their wave-functions are constant in x^μ ; the multi-branes BPS configurations are the bottom of a continuum spectrum of m^2 . A BPS bound state is instead a *normalizable* eigenstate of m^2 whose mass (or tension) saturates the appropriate BPS bound, $m^2 = |Z|^2$.

12.7.1 F1-D1 Bound States

We look for bound states of p F-strings²³ with q D-strings in IIB. All strings are at rest and aligned in the (periodic) 1-direction. These states have p units of winding F1-charge and q units of the RR D1-charge. For a state with these quantum numbers, the SUSY algebra (12.27)–(12.29) reduces to

$$\frac{1}{2} \left\{ \begin{bmatrix} Q_\alpha \\ \tilde{Q}_\alpha \end{bmatrix}, \begin{bmatrix} Q_\beta^\dagger & \tilde{Q}_\beta^\dagger \end{bmatrix} \right\} = M \delta_{\alpha\beta} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{L_1}{2\pi\alpha'} (\Gamma^0 \Gamma^1)_{\alpha\beta} \begin{bmatrix} p & q/g \\ q/g & -p \end{bmatrix}, \quad (12.139)$$

with L_1 the period of X^1 . The eigenvalues of $\Gamma^0 \Gamma^1$ are ± 1 , so those of the RHS are

$$M \pm L_1 \frac{(p^2 + q^2/g^2)^{1/2}}{2\pi\alpha'}. \quad (12.140)$$

The LHS of (12.139) is non-negative. This implies a BPS bound on the total mass per unit length

$$\frac{M}{L_1} \geq \frac{(p^2 + q^2/g^2)^{1/2}}{2\pi\alpha'} \stackrel{\text{def}}{=} \tau_{(p,q)}. \quad (12.141)$$

This inequality is saturated by the F-string which has $(p, q) = (1, 0)$ and by the D-string which has $(p, q) = (0, 1)$. A state which saturates the bound (12.141) is invariant under 16 SUSYs, i.e. it is half-BPS.

The sum of the energies per unit length of one F-string and one D-string is

²³ Here and below F-strings (or F1 brane) stands for the fundamental superstring.

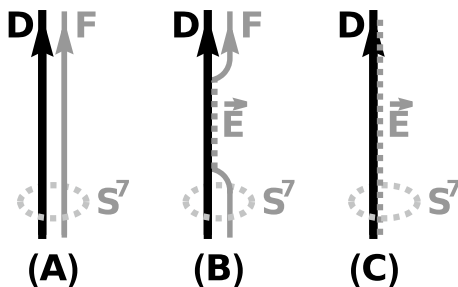


Fig. 12.2 (A) $\rightarrow$ (B): the F-string breaks with its endpoints ending in the D-string. (B) $\rightarrow$ (C): the endpoints move off to infinity, leaving behind the D-string with a non-trivial electric field $\vec{E}$. S^7 is a 7-sphere at infinity. The flux of $(*H)^{(7)}$ through S^7 measures (by Gauss law) the F-string conserved charge (“winding”). This conserved charge remains invariant in the process (A) $\rightarrow$ (C)

$$\tau_{(0,1)} + \tau_{(1,0)} = \frac{g^{-1} + 1}{2\pi\alpha'} > \frac{(g^{-2} + 1)^{1/2}}{2\pi\alpha'} \equiv \tau_{(1,1)}. \quad (12.142)$$

The BPS bound is not saturated, and hence this two-object state is *not* supersymmetric. In fact between a D- and F-string there is a non-zero attractive force. The presence of an *attractive* force suggests that a bound state may form.

Indeed the system can lower its energy through the process illustrated in **Fig. 12.2**. The F-string breaks and its two endpoints get attached to the D-string. The endpoints then can move to infinity along the D1, thus dissolving the F-string into the D-string. This cannot be the full story since the F-string carries the *conserved* NS-NS winding charge which is measured by the flux of $*H$ on the 7-sphere at infinity linked to the strings. Being conserved, this flux should be still non-zero in the final configuration. We describe the mechanism which ensures the conservation of the winding number in the process. The endpoint of an incoming (resp. outgoing) F-string carries a charge -1 (resp. $+1$) for the D1 world-sheet gauge field A_a . By 2d Gauss law a constant 2d electric field runs between the two F1 endpoints on the D1 world-sheet Ξ . This flux remains in the final configuration without the F-string. From the bulk plus D-string Born–Infeld action

$$S = S_{\text{bulk}} + S_{\text{BI},1}, \quad S_{\text{bulk}} = \frac{1}{2\kappa_0^2} \int d^{10}x \sqrt{-G} e^{-2\Phi} |dB|^2 + \dots \quad (12.143)$$

$$S_{\text{BI},1} = -T_1 \int_{\Xi} d^2\xi e^{-\Phi} [\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab})]^{1/2}, \quad (12.144)$$

we see that the equation of motion of the 2-form gauge field B ,

$$0 = \frac{\delta S}{\delta B(x)} = \frac{1}{\kappa_0^2} d(e^{-2\Phi} * dB(x)) + \overbrace{\frac{\delta S_{\text{BI},1}}{\delta B(x)}}^{\text{source}}, \quad (12.145)$$

contains a source term localized on the D1-brane world-sheet Ξ whose strength is proportional to the invariant electric flux on the D-string

$$\mathcal{F}_{ab} = F_{ab} + \frac{B_{ab}}{2\pi\alpha'}, \quad (12.146)$$

see **BOX 12.6** for the precise expression of the source term. Thus a D-string with a constant electric field on its world-sheet is an electric source for the NS gauge field $B_{\mu\nu}$, in other words it carries a non-zero F1 winding number. Both winding number and electric flux are integrally quantized, and p units of electric field produce p units of winding number (for a detailed argument see final remarks in **BOX 12.7**).

BOX 12.6 - Equation (12.146)

In 2d

$$[\det(G_{ab} + B_{ab} + 2\pi\alpha' F_{ab})]^{1/2} \equiv \sqrt{G} \left(1 + \frac{1}{2}(2\pi\alpha')^2 \mathcal{F}^{ab} \mathcal{F}_{ab}\right)^{1/2},$$

so that

$$\frac{\delta S_1}{\delta B^{\mu\nu}} = -\pi\alpha' T_1 e^{-\Phi} \left(1 + \frac{1}{2}(2\pi\alpha')^2 \mathcal{F}^{ab} \mathcal{F}_{ab}\right)^{-1/2} \mathcal{F}_{\mu\nu} \delta_{D1},$$

where δ_{D1} is a delta-function with support on the D-string world-sheet.

We claim that the final state (C) in **Fig. 12.2**, with charges $(p, q) = (1, 1)$, is half-BPS, i.e. preserves 16 supercharges. More precisely: *the F1-D1 system forms exactly one short supermultiplet of $\frac{1}{2}$ -BPS states (2^8 BPS states).*

We give two proofs: first by T -duality, and then by direct computation of the tension of the final state showing that it saturates the BPS bound.

First Proof: the T -Dual Point of View

The simplest proof that the bound state is BPS, is via T -duality in the 1-direction. The D1-brane becomes a D0-brane. The constant electric field is T -dual to a velocity

$$\dot{A}_1 \rightsquigarrow \frac{1}{2\pi\alpha'} \dot{X}_1, \quad (12.147)$$

so the T -dual state of a D1-brane with an electric flux is a D0-brane moving with constant velocity. The quantization of electric flux corresponds to quantization of momentum along the dual 1-direction which is compactified on a small circle of length $L'_1 = 4\pi^2\alpha'/L_1 \lll \sqrt{\alpha'}$. This state is invariant under the same number of supersymmetries as a D0-brane at rest, indeed under supercharges which are the Lorentz boost of the ones preserved by a D0 at rest. Since the number of unbroken SUSYs is invariant under T -duality, we conclude that the D1-F1 bound state is $\frac{1}{2}$ -BPS. The number of BPS states is also a T -duality invariant, and we conclude that the BPS states form a *single* ultrashort supermultiplet (2^8 states).

More generally the T -duality argument applies to bound states of one D1 with any number $p \in \mathbb{Z}$ of F1's: on the world-sheet we have p units of 2d electric flux, and the T -dual state is a D0-particle with p units of compact momentum, i.e. $P_1 = 2\pi p/L'_1$. Such states form a single $\frac{1}{2}$ -BPS supermultiplet.

BOX 12.7 - The BPS bound is saturated

In 2d Lorentzian signature $(+, -)$ and flat target space, setting $f \equiv 2\pi\alpha' \mathcal{F}_{01}$, one has

$$G_{ab} + 2\pi\alpha' \mathcal{F}_{ab} = \begin{pmatrix} 1 & f \\ -f & -1 \end{pmatrix}_{ab}$$

and then the world-sheet Born–Infeld Lagrangian reads

$$L = -\tau_{(0,1)} \left[-\det(G_{ab} + 2\pi\alpha' \mathcal{F}_{ab}) \right]^{1/2} \equiv -\tau_{(0,1)} (1 - f^2)^{1/2}$$

. The momentum conjugate to the D-string gauge field A_1 is

$$p = \frac{\partial L}{\partial \mathcal{F}_{0,1}} \equiv 2\pi\alpha' \frac{\partial L}{\partial f} = -2\pi\alpha' \tau_{(0,1)} \frac{f}{(1 - f^2)^{1/2}}$$

. Inverting this relation, one gets f as a function of p :

$$f(p)^2 = \frac{p^2}{(2\pi\alpha' \tau_{(0,1)})^2 + p^2} \equiv \frac{p^2}{g^{-2} + p^2},$$

since $2\pi\alpha' \tau_{(0,1)} \equiv g^{-1}$. The world-sheet Hamiltonian is

$$\begin{aligned} H &= p \cdot \frac{f(p)}{2\pi\alpha'} - L(f(p)) = \frac{1}{2\pi\alpha'} \left[\frac{p^2}{(g^{-2} + p^2)^{1/2}} + \frac{1}{g} \left(1 - \frac{p^2}{g^{-2} + p^2} \right)^{1/2} \right] = \\ &= \frac{1}{2\pi\alpha'} (p^2 + g^{-2})^{1/2}. \end{aligned}$$

For a clean computation, one should wrap the D-string on a S^1 ; then the conjugate variable p is quantized in integral units. This is T -dual to the quantization of the usual momentum for the D0-brane, and hence corresponds to the quantization of the 2-form string charge, i.e. the closed string winding number; from the D-string world-sheet gauge field viewpoint this follows from the characters of the $U(1)$ gauge group (that is, from the quantization of the electric charge). The integer p is just the number of fundamental strings dissolved in our D-string. The eigenvalue of H then precisely saturates the inequality (12.141) with $q = 1$ and arbitrary p .

Second Proof: Direct Computation of the Tension

The detailed direct computation in BOX 12.7 shows that one D-string and an arbitrary number p of F-strings form a bound state whose tension saturates the BPS bound. At weak coupling, $g \ll 1$, the BPS F1-D1 bound state is *maximally bound*, indeed its binding tension

$$\tau_{(1,0)} + \tau_{(0,1)} - \tau_{(1,1)} = \frac{1}{2\pi\alpha'g} \left[1 + g - (1 + g^2)^{1/2} \right] = \frac{1}{2\pi\alpha'} - O(g), \quad (12.148)$$

is almost equal to the entire F-string tension.

String World-Sheet Description

The final state, a D1 with a *constant* electric flux on its world-sheet, has a simple description in terms of 2d QFT theory on the world-sheet Σ of the F-strings ending on the D1. In this description, the non-zero electric field may enter only through the boundary condition at the string endpoint on the D1-brane. The variation of the world-sheet action now takes the form

$$\begin{aligned} \delta \left(S_{\text{bulk}} + i \oint_{\partial\Sigma} ds A_v(X) \partial_s X^v \right) = \\ = -\frac{1}{2\pi\alpha'} \int_{\Sigma} \delta X^\mu \partial^\alpha \partial_\alpha X_\mu + \oint_{\partial\Sigma} ds \delta X^\mu \left(\frac{1}{2\pi\alpha'} \partial_n X_\mu + i F_{\mu\nu} \partial_s X^\nu \right), \end{aligned} \quad (12.149)$$

where we used

$$\begin{aligned} i \delta \left(\oint_{\partial\Sigma} ds A_v(X) \partial_s X^v \right) &= i \oint_{\partial\Sigma} ds \left[\delta X^\mu \partial_\mu A_v \partial_s X^v + A_\mu (\partial_s \delta X^\mu) \right] = \\ &= i \oint_{\partial\Sigma} ds \left[\delta X^\mu \partial_\mu A_v \partial_s X^v - \delta X^\mu (\partial_v A_\mu) \partial_s X^v \right] \equiv i \oint_{\partial\Sigma} ds \delta X^\mu F_{\mu\nu} \partial_s X^\nu. \end{aligned} \quad (12.150)$$

Equation (12.149) implies the linear boundary condition

$$\partial_n X_\mu + 2\pi i \alpha' F_{\mu\nu} \partial_s X^\nu = 0. \quad (12.151)$$

The winding charge of this state in the 1-direction, $\int \partial_n X_1 / L_1$, is proportional to $F_{1,0}$, i.e. to the electric flux on the D1, in full agreement with the result we got from the QFT living on the D1 world-sheet.

Generalization to (p, q) Strings

We have seen that p F-strings and one D-string form *one* $\frac{1}{2}$ -BPS supermultiplet of $(p, 1)$ bound states. We wish to study the general case of p F1's and q D1's and understand for which pairs (p, q) there are BPS bound states and how many. This requires to extend the computation of the tension in **BOX 12.7** to $q \geq 2$. The story now gets more involuted because for $q \geq 2$ the effective theory on the D1 world-sheet is a non-Abelian Yang–Mills theory with gauge group $U(q)$. In 2d the YM coupling has units of mass-squared, so is a *relevant* coupling, and the D1 world-sheet QFT gets *strongly coupled* in the IR: the existence of BPS bound states becomes a fully non-perturbative question when $q \geq 2$. Luckily, the world-sheet QFT is supersymmetric, and we have powerful non-perturbative techniques to address the question. The non-perturbative analysis leads to the following.

Claim 12.1 *There is precisely **one** $\frac{1}{2}$ -BPS supermultiplet of bound states of p F1's and q D1's for all pairs of integers (p, q) with p, q **relatively prime**, $\gcd(p, q) = 1$.*

The bound state of p F1's and q D1's is a (p, q) -string²⁴. When p, q are *not* coprime,

$$(p, q) = (kp', kq'), \quad k > 1, \quad \gcd(p', q') = 1, \quad (12.152)$$

the system is marginally unstable against breaking into k (p', q') -strings: indeed a set of k parallel (p', q') -strings at rest is still a BPS configuration, just as a configuration of k parallel D1-strings which is the special case $(p', q') = (0, 1)$.

Let us give the physical intuition beyond the **Claim**.

Claim 12.1: Physical Heuristics The world-sheet theory is a $U(1) \times SU(q)$ gauge theory. The non-Abelian degrees of freedom become strongly coupled in the IR, while the $U(1)$ ones remain IR free. The string endpoints now have charge $\pm 1/q \bmod 1$ for the overall $U(1)$, see Eq. (12.156). Since the non-Abelian d.o.f. are confined, they are not part of the low-energy description, and the IR effective theory is purely Abelian. We are reduced back to the computation in **BOX 12.7** with two modifications: (i) the tension of q D1's is q times the tension of a single D-string, and (ii) the electric flux is now quantized in integral multiples of $1/q$ times the original quantum.²⁵ Modification (i) multiplies the Hamiltonian in **BOX 12.7** by the overall factor q , while (ii) replaces p units of flux with p/q units ($p \in \mathbb{Z}$). Then

$$H = \frac{q}{2\pi\alpha'} \left(\frac{p^2}{q^2} + \frac{1}{g^2} \right)^{1/2} \equiv \frac{1}{2\pi\alpha'} \left(p^2 + \frac{q^2}{g^2} \right)^{1/2}, \quad (12.153)$$

and the BPS bound is saturated. If (p, q) are coprime, the BPS string cannot fall apart into pieces carrying smaller charges because of the energy barrier implied by the BPS bound.

This intuitive argument can be fully justified making it into a non-perturbative proof. The world-sheet QFT is supersymmetric: then, by the usual Witten-index arguments [31], as long as we are *only* interested in establishing (at the full non-perturbative level) the existence of BPS states belonging to ultrashort SUSY supermultiplets, we can use *any* approximate description which gets exact in some asymptotic limit, provided it gives an unambiguous answer. Our heuristic description fulfills these criteria for p, q coprime. In the next paragraph, we present a more detailed and rigorous non-perturbative proof of the **Claim** following Witten [29].

★ **Claim 12.1: Detailed & Rigorous Proof**

We consider the problem of *counting* the number $N(p, q)$ of half-BPS supermultiplets of bound states in the system of p F1's and q D1's from the viewpoint of the QFT $\mathcal{T}_q$ which lives on the world-sheet of the q D1 strings. The previous “heuristic” argument is a cartoon of the rigorous story

²⁴ Not to be confused with a p - p' string which is a string with endpoints on Dp - Dp' -branes.

²⁵ Besides the 't Hooft argument leading to Eq. (12.156), the quantization of the electric flux in $1/q$ units can be understood geometrically in string theory: see *Note 12.7* below.

[29]. $\mathcal{T}_q$ is a 2d SUSY gauge model with 16 supercharges, $Spin(8)$ R-symmetry,²⁶ and gauge group $U(q)$. The field content is a gauge vector A_μ , 8 scalars X^i , and 16 MW fermions, all in the adjoint of the gauge group $U(q)$.

The Hilbert space of $\mathcal{T}_q$ decomposes into sectors of definite F1-charge p

$$\mathbf{H}_{\mathcal{T}_q} = \bigoplus_p \mathbf{H}_{(p, q)}. \quad (12.154)$$

Let us recall how the sectors $\mathbf{H}_{(p, q)}$ arise. The gauge group is

$$U(q) \simeq [U(1) \times SU(q)]/\mathbb{Z}_q. \quad (12.155)$$

As reviewed in **BOX 12.8**, in a 2d gauge theory whose fields are all in the adjoint representation, the sectors of the Hilbert space with definite gauge charges are labeled by elements of the *dual group* of the center of the gauge group; in our case

$$\text{group of gauge charges} \equiv \text{Hom}(Z(U(n)), U(1)) \simeq \left\{ \left(\frac{p}{q}, p \right) \in \frac{1}{q}\mathbb{Z} \oplus \mathbb{Z}_q \right\}, \quad (12.156)$$

where the projection of the charge in the summand $\frac{1}{q}\mathbb{Z}$ is the Abelian $U(1)$ charge while the projection on $\mathbb{Z}_q$ is the $SU(q)$ charge—we stress that *the $SU(q)$ charge is defined only mod q* .

The degrees of freedom of $\mathcal{T}_q$ separate into Abelian and non-Abelian ones. From the point of view of the system of strings, the Abelian d.o.f. describe the overall motion of the system's center of mass, while the non-Abelian ones are the internal dynamics of the system which is described by some $SU(q)$ SUSY gauge theory $\mathbf{Th}_q$ invariant under 16 supercharges. While the $\mathbf{Th}_q$ Lagrangian is rather complicated, at low-energy it is well approximated by its truncation to two derivatives, that is, by the standard 2d 16-supercharges super-Yang–Mills with gauge group $SU(q)$, i.e. the 2d dimensional reduction of 4d $\mathcal{N} = 4$ SYM. We shall work with this low-energy QFT, our conclusions being guaranteed to be exact by standard arguments about SUSY protected quantities. Thus

$$\mathbf{H}_{(p, q)} = \mathbf{H}_{(p, q)}^{\text{Ab}} \otimes \mathbf{H}_{(p \bmod q, q)}^{\text{n-Ab}}. \quad (12.157)$$

The half-BPS states of the (p, q) system of strings are linear combinations of vectors

$$|\psi^{\text{Ab}}\rangle \otimes |\psi^{\text{n-Ab}}\rangle \in \mathbf{H}_{(p, q)} \quad (12.158)$$

which are invariant under 16 supercharges, that is, under *all* supersymmetries of $\mathcal{T}_q$. A state invariant under all SUSYs is a *SUSY vacuum*, so both states $|\psi^{\text{Ab}}\rangle$ and $|\psi^{\text{n-Ab}}\rangle$ are SUSY vacua of the Abelian and non-Abelian theories, respectively. A half-BPS *bound state* of the (p, q) system thus corresponds to a *normalizable* SUSY-preserving vacuum $|\psi^{\text{n-Ab}}\rangle$ of the non-Abelian theory $\mathbf{Th}_q$ which belongs to the Hilbert space sector $\mathbf{H}_{(p \bmod q, q)}^{\text{n-Ab}}$.

The Abelian Sector The Abelian part of the story is weakly coupled in the IR. Its 16 fermions are IR free, and their zero-modes generate a Clifford algebra in $k = 16$ dimensions, hence the Abelian vacua form a $\text{Cl}(16)$ Clifford module, and we have 2^8 vacuum states (2^7 bosonic and 2^7 fermionic). This is just the number of states in a half-BPS supermultiplet in 10d IIB.

The bosonic sector of the Abelian theory works as in **BOX 12.7**, except that the tension of q D1-branes is $q\tau_{(0,1)}$ instead of $\tau_{(0,1)}$

$$\tau_{(0,1)} \equiv \frac{1}{2\pi\alpha'} \frac{1}{g} \rightarrow \frac{1}{2\pi\alpha'} \frac{q}{g}, \quad (12.159)$$

²⁶ The R-symmetry is *chiral* in the sense that the left-moving fermions are in the $\mathbf{8}_s$ of $Spin(8)$ while the right-movers are in the *inequivalent* representation $\mathbf{8}_c$. This symmetry is nevertheless free of 't Hooft anomalies, since the quadratic Casimirs of the two representations are equal.

BOX 12.8 -'t Hooft twisted boundary conditions and all that

Compare with 't Hooft analysis [32, 33] of non-perturbative 4d Yang–Mills theory

Topological sectors in 2d gauge theory We consider the topological sectors of a 2d YM theory, with gauge group G , coupled to matter in the *adjoint* representation, quantized in an Euclidean torus $S^1 \times S^1$ seen as a rectangle with opposite sides identified. We write L, β for the sides of the rectangle ($\beta \equiv$ inverse temperature) and Φ for *all* fields including A_μ . Notice that the gauge group which acts effectively on the *local* d.o.f. is $G/Z(G)$, where $Z(G)$ is the center of G . The fields $\Phi(x_1, x_2)$ are periodic up to a gauge transformations $\Omega_1(x_2), \Omega_2(x_1) \in G$

$$\Phi(x_1 + L, x_2) = \text{Ad}_{\Omega_1(x_2)} \Phi(x_1, x_2), \quad \Phi(x_1, x_2 + \beta) = \text{Ad}_{\Omega_2(x_1)} \Phi(x_1, x_2) \quad . \quad (\spadesuit)$$

Consistency at the corners requires

$$\text{Ad}_{\Omega_1(x_2 + \beta)} \text{Ad}_{\Omega_2(x_1)} = \text{Ad}_{\Omega_2(x_1 + L)} \text{Ad}_{\Omega_1(x_2)}$$

that is,

$$\Omega_1(x_2 + \beta) \Omega_2(x_1) = \zeta \Omega_2(x_1 + L) \Omega_1(x_2), \quad \zeta \in Z(G) \quad . \quad (\clubsuit)$$

The central element ζ does not depend on the gauge, so is a topological invariant of the configuration: the topological sectors in the path integral are labeled by elements of the center $Z(G) \subset G$. The Hilbert space charged sectors are then classified by the dual group $\text{Hom}(Z(G), U(1))$

$$H = \bigoplus_{\xi \in \text{Hom}(Z(G), U(1))} H_\xi.$$

The trace on the Hilbert sector H_ξ is

$$\text{Tr}_\xi [(-1)^{\alpha F} e^{-\beta H}] = \sum_{\zeta \in Z(G)} \xi(\zeta) Z^\alpha(L, \beta; \zeta), \quad \alpha = 0, 1$$

where for $\alpha = 1$ $Z^\alpha(L, \beta; \zeta)$ is the path integral over fields in the topological sector $\zeta \in Z(G)$, while for $\alpha = 0$ the fermions pick up an extra minus sign when going around the time circle.

Examples: (1) $G = U(1)$: $\text{Hom}(U(1), U(1)) \cong \mathbb{Z}$, the electric charge which is integrally quantized. (2) $G = SU(q)$: $\text{Hom}(Z(SU(q)), U(1)) \simeq \mathbb{Z}_q \equiv \mathbb{Z}/q\mathbb{Z}$. For $e \in \mathbb{Z}_q$ one has

$$\text{Tr}_e [(-1)^{\alpha F} e^{-\beta H}] = \sum_{k=0}^{q-1} e^{2\pi i k e / q} Z^\alpha(L, \beta; e^{2\pi i k / q}), \quad e \in \mathbb{Z}/q\mathbb{Z}, \quad (\heartsuit \diamondsuit)$$

Note that the physics depends on e only mod q .

and the net effect in the final formula of **BOX 12.7** is that g gets replaced by g/q , yielding an Abelian contribution to the tension

$$\frac{1}{2\pi\alpha'} (p^2 + q^2/g^2)^{1/2} \quad (12.160)$$

which already saturates the BPS tension (cf. the “heuristic” argument). We conclude that, to have a BPS state, the contribution to the energy from the non-Abelian sector should vanish, as we expect since $|\psi^{\text{n-Ab}}\rangle$ should be a SUSY vacuum of the internal non-Abelian theory $\mathbf{Th}_q$.

The Internal Theory $\mathbf{Th}_q$ The number $N(p, q)$ of half-BPS supermultiplets of (p, q) bound states is just the number of SUSY preserving, *normalizable* vacua in the Hilbert sector $\mathbf{H}_{(p \bmod q, q)}^{n\text{-Ab}}$ of the internal QFT $\mathbf{Th}_q$. In this theory p is defined mod q , so we have

$$N(p + q, q) = N(p, q), \quad (12.161)$$

a fact which has a deep physical explanation (see Chap. 13). In a normalizable vacuum $|\psi^{n\text{-Ab}}\rangle$, the R-symmetry $Spin(8)$ is unbroken by Coleman's theorem [2]. Since the fermions of $\mathbf{Th}_q$ transform in a spinorial representation of $Spin(8)$, $(-1)^F$ can be identified with the 2π rotation in $Spin(8)$, so

$$(-1)^F |\psi^{n\text{-Ab}}\rangle = |\psi^{n\text{-Ab}}\rangle, \quad (12.162)$$

and all normalizable vacua are bosonic (so the index coincides with the total number of vacua).

Normalizable versus Non-normalizable Vacua The 2d non-Abelian theory $\mathbf{Th}_q$ has also *non-normalizable* SUSY vacua. They arise in sectors (p, q) with $\gcd(p, q) = k > 1$ (in particular $p = 0$ and any $q > 1$). Such vacua are *non-gapped*, hence they do not represent half-BPS bound states. The existence of non-normalizable $\mathbf{Th}_q$ vacua in sectors with (p, q) not coprime corresponds to the physical fact that we may split the (p, q) configuration, with no cost in energy, into k parallel $(p/k, q/k)$ -strings. Such a static configuration is BPS for all separations between the parallel $(p/k, q/k)$ -strings but represents a SUSY ground state of $\mathbf{Th}_q$ iff the k $(p/k, q/k)$ -strings are in a zero-eigenstate of their relative momenta, and this state cannot be normalizable since the spectrum of momentum is continuous. These non-normalizable states represent multi-brane BPS configurations. *Normalizable* SUSY vacua are quite different:

Claim 12.2 *In a supersymmetric 2d QFT with an exact R-symmetry $SO(N)$ and $N \geq 5$, all non-normalizable vacua which preserve SUSY are gapped.*

Proof Suppose the vacuum is not gapped. The IR limit is then a *non-trivial* SCFT invariant under the full set of supercharges which are rotated by an *unbroken* R-symmetry $SO(N)$ (by Coleman's theorem). There is no such SCFT by the classification in Sect. 2.10.1. Contradiction!

Let us check that, in a (p, q) sector with p, q coprime, if a SUSY preserving state exists, it is automatically normalizable, and explain why this fails when p, q are *not* coprime. From the viewpoint of the 2d YM theory, the statement is that to send the scalars X^i to infinity in any direction we need to pay an amount of energy larger than a certain positive value ΔE when (p, q) are coprime and zero otherwise. In other words, there must be a quantum effective potential $\mathcal{V}(X^i)$ whose value at infinity is $\geq \Delta E$, so that a zero-energy eigenstate (a SUSY vacuum) has a wave-function which vanishes exponentially at infinity in field space. We exploit the fact, obvious from the description of the YM charge sectors, that the Hilbert sector $\mathbf{H}_{(p \bmod q, q)}^{n\text{-Ab}}$ is obtained by inserting at infinity a “particle” in any representation $\mathcal{R}$ of $SU(q)$ on which the center $Z(SU(q)) \simeq \mathbb{Z}_q$ acts by the character $k \mapsto e^{2\pi i k p/q}$ [34]; e.g. we may take $\mathcal{R} = \wedge^p F$ with F the fundamental of $SU(q)$.

Now we try to make X^i large while keeping the energy of the state zero. Since the scalar potential has the form

$$\propto \sum_{i,j} \text{tr}([X^i, X^j][X^i, X^j]), \quad (12.163)$$

for large X^i , the matrices X^i should all commute, hence they may be diagonalized simultaneously. In the large X^i regime the non-Abelian d.o.f. are weakly coupled, so the semi-classical analysis is reliable. The large v.e.v.'s of the scalars all take value in a Cartan subalgebra of $\mathfrak{su}(q)$, and the gauge group breaks to a group of the form (for generic large values of X^i one has $G' = U(1)^{q-1}$)

$$G' = \prod_{i=1}^{s+1} U(k_i) / U(1), \quad \sum_{i=1}^{s+1} k_i = q, \quad s \geq 1, \quad 1 \leq k_i < q, \quad (12.164)$$

so we have a non-trivial unbroken Abelian gauge symmetry $U(1)^s$ whose generators in the fundamental representation, Q_a ($a = 1, \dots, s$), are (proportional to) diagonal matrices with k_a eigenvalues $(q - k_a)$ and $(q - k_a)$ eigenvalues $-k_a$. The values of the Q_a charge for the various components of the representation $\wedge^p F$ are

$$\left\{ s(q - k_a) - (p - s)k_a \equiv sq - pk_a, \quad s = 0, 1, \dots, p \right\} \quad (12.165)$$

hence all congruent to $-pk_a \bmod q$; since $k_a < q$, if p is coprime to q no component of the representation $\wedge^p F$ at infinity has vanishing charge for the unbroken $U(1)$ gauge symmetries, hence in this (p, q) sector the background light Abelian electric flux cannot be zero (by Gauss law) and thus for large X^i the energy must be strictly positive, as was to be shown. On the contrary, assume $(p, q) = (kp', kq')$ ($k > 1$) and take the X^i to infinity along a direction which breaks

$$SU(q) \rightarrow G' = U(q')^k / U(1), \quad (12.166)$$

while the fundamental F of $SU(q)$ decomposes in representation of G' as

$$F = \bigoplus_{i=1}^k F_i, \quad F_i \text{ the fundamental of the } i\text{-th factor } U(q'). \quad (12.167)$$

Consider the subrepresentation

$$N := \bigotimes_{i=1}^k (\wedge^{p'} F_i) \subset \wedge^p F. \quad (12.168)$$

All elements of N are neutral for *all* unbroken Abelian gauge symmetries generated by the Q_a . Hence putting at infinity a $\wedge^p F$ “particle” in a color state belonging to N we construct low-energy states in the (p, q) sector without background electric fields, whose energy remains zero. Thus we may send X^i to infinity in this particular direction at no cost in energy. Note that this direction in field space corresponds to splitting the (p, q) system into k (p', q') systems and then taking their separation to be large. So the non-normalizability of the state from the viewpoint of the 2d YM precisely matches the physical idea of a stack of parallel (p', q') -strings which, being BPS for arbitrary separations, at zero momentum is a non-normalizable SUSY-preserving vacuum.

The bottom line is that for (p, q) coprime the SUSY preserving vacua of $\mathbf{Th}_q$, if present *at all*, should be normalizable, and hence gapped, by **Claim 12.2**. In particular, for (p, q) coprime *all* SUSY vacua of $\mathbf{Th}_q$ describe a *bona fide* bound state of p F1’s and q D1’s.

It remains to compute the number $N(p, q)$ of such SUSY preserving vacua for given coprime p, q . We already know that $N(p, q)$ depends on p only mod q .

Computing $N(p, q)$ for $\gcd(p, q) = 1$ To count the gapped SUSY vacua of $\mathbf{Th}_q$ we may follow the standard deformation strategy: a gapped SUSY vacuum remains SUSY and gapped for all sufficiently small perturbation which preserves *some* supersymmetry.

It is convenient to write the Lagrangian of the 2d $(8, 8)$ SYM $\mathbf{Th}_q$ in terms of $(2, 2)$ superfields. The 2d $(8, 8)$ vector supermultiplet decomposes into a $(2, 2)$ vector superfield (whose bosonic fields are a vector A_μ and a complex scalar σ) and 3 chiral superfields Φ_i ($i = 1, 2, 3$), all in the adjoint of the gauge group $SU(q)$. To see this, recall that the 2d $(8, 8)$ SYM is obtained by dimensional reduction to 2d of 10d SYM. The reduction can be done in two steps: first $10d \rightarrow 4d$ and then $4d \rightarrow 2d$. In 4d we get the $\mathcal{N} = 4$ SYM theory which, written in terms of $\mathcal{N} = 1$ superfields, consists of a gauge supermultiplet and 3 chiral superfields in the adjoint. Written in terms of $\mathcal{N} = 1$ superfields the 4d $\mathcal{N} = 4$ theory has Lagrangian

$$L = i\tau \int d^2\theta \operatorname{tr} \left(W_\alpha W^\alpha + \Phi_1 [\Phi_2, \Phi_3] \right) + \text{h.c.} \quad (12.169)$$

while the 2d (8, 8) Lagrangian is obtained from this one by taking all fields to be independent of x^2, x^3 and setting $\sigma = A_2 + iA_3$. The 2d scalar potential has the form

$$\begin{aligned} & \sum_{i,j} \text{tr}([\Phi_i, \Phi_j][\Phi_i, \Phi_j]^\dagger) + \text{tr}([\sigma, \bar{\sigma}]^2) + \\ & + \sum_i \text{tr}([\Phi_i, \sigma][\Phi_i, \sigma]^\dagger + [\Phi_i, \bar{\sigma}][\Phi_i, \bar{\sigma}]^\dagger), \end{aligned} \quad (12.170)$$

so that in a zero-energy vacuum all fields $\Phi_i, \bar{\Phi}_i, \sigma$ and $\bar{\sigma}$ are simultaneously diagonalizable: generically they take value in a Cartan subalgebra $\mathfrak{u}(1)^{\oplus(q-1)} \subset \mathfrak{su}(q)$.

A priori, understanding the vacua of $\mathbf{Th}_q$ is hard since the model is strongly coupled in the IR. We look for a deformation of the theory which simplifies the counting of SUSY vacua. It is technically convenient to deform $\mathbf{Th}_q$ while preserving 4 out of the 16 supercharges. This is done by deforming the superpotential in (12.169) as

$$W(\Phi) \equiv \text{tr}(\Phi_1[\Phi_2, \Phi_3]) \longrightarrow \text{tr}(\Phi_1[\Phi_2, \Phi_3]) + \frac{1}{2} m \sum_i \text{tr}(\Phi_i \Phi_i). \quad (12.171)$$

On general grounds, a gapped SUSY vacuum of the unperturbed theory $\mathbf{Th}_q$ remains SUSY and gapped for all sufficiently small m . It is intuitively clear that taking the mass m to be large can only improve the mass gap, so increasing $|m|$ will not close the gap, and our gapped, normalizable SUSY vacua will remain gapped, normalizable, and supersymmetric for *all* values of the deformation parameter m . On a more technical level, the field redefinitions $\Phi_i \rightarrow \Phi_i/\lambda$ and $\theta \rightarrow \lambda^{3/2} \theta$ make $m \rightarrow \lambda m$ leaving the Lagrangian invariant up to a continuous deformation of the D -terms, which are always IR irrelevant in a 2d (2, 2) QFT: hence taking m large cannot destroy the energy gap.

The deformation argument works also in the reverse direction. Suppose that for large m we find a normalizable, gapped, and supersymmetric vacuum. Using the above rescaling argument, we see that continuously deforming m down to some small non-zero value ϵ the state remains a normalizable, gapped, SUSY vacuum. Now taking $\epsilon \rightarrow 0$ we get a zero-energy state which a priori may be non-normalizable and/or non-gapped. However, we already know that in a (p, q) sector with $\gcd(p, q) = 1$ a zero-energy state has a normalizable wave-function, and also that a normalizable zero-energy state cannot break $S O(8)$ (by Coleman theorem) and hence is gapped by **Claim 12.2**:

Fact 12.2 The number $N(p, q)$ of normalizable, gapped, SUSY vacua of the non-Abelian theory $\mathbf{Th}_q$ in a sector $\mathbf{H}_{(p \bmod q, q)}^{\text{n-Ab}}$ with $\gcd(p, q) = 1$ is equal to the number of normalizable, gapped, SUSY vacua of the $m \rightarrow \infty$ limit of the deformed 2d (2, 2) theory (12.171) in the sector $\mathbf{H}_{(p \bmod q, q)}^{\text{n-Ab}}$.

A SUSY vacuum of the deformed theory satisfies the condition

$$dW \equiv d\Phi_i \frac{\partial W(\Phi)}{\partial \Phi_i} = 0, \quad (12.172)$$

which, for the superpotential (12.171), becomes

$$[\Phi_i, \Phi_j] = m \epsilon_{ijk} \Phi_k, \quad (12.173)$$

that is, when evaluated on a SUSY vacuum the $q \times q$ traceless matrices $L_i \equiv \Phi_i/m$ yield a representation of $\mathfrak{sl}(2)$. Then, modulo gauge equivalence, the solutions to $dW = 0$ are in one-to-one correspondence with isoclasses of degree q representations V of $\mathfrak{sl}(2)$. One has

$$V = \bigoplus_{\ell} V_{\ell}^{\oplus n_{\ell}}, \quad \sum_{\ell} (2\ell + 1) n_{\ell} = q, \quad (12.174)$$

where V_ℓ is the irreducible $\mathfrak{sl}(2)$ representation of spin ℓ . We set

$$s + 1 \stackrel{\text{def}}{=} \#\{\ell: n_\ell \neq 0\}. \quad (12.175)$$

For large m the vacuum associated with a representation V as in (12.174) breaks the gauge group to

$$G \rightarrow U(1)^s \times \prod_{\ell} SU(n_\ell). \quad (12.176)$$

When $s \geq 1$ we have an unbroken Abelian gauge group. Hence, by the argument around Eq. (12.165), in a (p, q) sector with $\gcd(p, q) = 1$ a low-lying state with $s \geq 1$ has a strictly positive energy in the limit $m \rightarrow \infty$, and hence cannot arise from the deformation of a normalizable SUSY state.

In the $s = 0$ case we have only one non-zero $n_\ell \equiv n$, for some spin ℓ , and an unbroken $SU(n)$ gauge group with $n \equiv q/d$ where $d \equiv 2\ell + 1$ is a divisor of q . The three adjoint scalars Φ_i get a large mass for $m \rightarrow \infty$ and decouple. We remain with a 2d $(2, 2)$ SYM with gauge group $SU(n)$. For $n \geq 2$ this is a strongly coupled QFT. This theory cannot have a mass gap: indeed, it has a chiral $U(1)$ R-symmetry with a non-zero 't Hooft anomaly. By the usual anomaly matching argument [35], the theory cannot be trivial in the infra-red. Therefore $s = 0$, $n \geq 2$ vacua cannot arise from a deformation of normalizable, gapped, SUSY preserving vacua of $\mathbf{Th}_q$.

We remain with $s = 0$ and $n = 1$. The gauge group is totally broken and *all* fields get a large mass. This vacuum is the only normalizable, gapped, SUSY-preserving state in the system. We conclude that

$$N(p, q) = 1 \quad \text{for } \gcd(p, q) = 1. \quad (12.177)$$

Claim 12.1: Evidence from Effective SUGRA We saw in Chap. 8 that the e.o.m. of the low-energy Type IIB effective theory are invariant under a $SL(2, \mathbb{Z})$ duality group, and have half-BPS solutions corresponding to the supergravity fields sourced, respectively, by a F-string and D-string. Hence IIA SUGRA essentially predicts the existence of $(1, 0)$ - and $(0, 1)$ -strings. The group $SL(2, \mathbb{Z})$ acts on the 2-form fluxes, hence of the string charges (p, q) , as

$$\begin{pmatrix} q \\ p \end{pmatrix} \rightarrow \begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z}), \quad (12.178)$$

and the $SL(2, \mathbb{Z})$ orbit of the F-string charge $(1, 0)$ is the set of all *coprime* pairs (p, q) . Hence, acting with the $SL(2, \mathbb{Z})$ symmetry on the explicit BPS solution representing a F-string, we get BPS solutions which represent (p, q) -strings for all coprime (p, q) . Thus SUGRA provides further evidence for **Claim 12.1**.

12.7.2 D0-Dp Bound States

By T -duality the study of bound states of Dp - Dp' branes can be reduced to the case of $p' = 0$ and we focus on a Type IIA system with the charges of one D0-particle and one Dp -brane wrapped on the large torus T^p

$$T^p = \{(x^1, \dots, x^p) \sim (x^1 + n_1 L_1, \dots, x^p + n_p L_p), \quad n_a \in \mathbb{Z}\} \quad (12.179)$$

of volume $V_p = L_1 L_2 \cdots L_p$ ($L_a \gg \sqrt{\alpha'}$). The SUSY algebra takes the form

$$\frac{1}{2} \left\{ \begin{bmatrix} Q_\alpha \\ \tilde{Q}_\alpha \end{bmatrix}, \begin{bmatrix} Q_\beta^\dagger & \tilde{Q}_\beta^\dagger \end{bmatrix} \right\} = M \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \delta_{\alpha\beta} + \begin{bmatrix} 0 & Z \\ -Z^\dagger & 0 \end{bmatrix} \Gamma_{\alpha\beta}^0 \quad (12.180)$$

where M is the mass and

$$Z = \tau_0 + \tau_p V_p \beta, \quad \text{and} \quad \beta = \beta^1 \cdots \beta^p. \quad (12.181)$$

The mass of the Dp -brane (seen as a particle in $(9-p)$ d) is $M_p = \tau_p V_p$.

The non-negativity of the LHS of (12.180) implies that

$$M^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \geq \begin{bmatrix} 0 & Z \\ -Z^\dagger & 0 \end{bmatrix} \Gamma^0 \begin{bmatrix} 0 & Z \\ -Z^\dagger & 0 \end{bmatrix} \Gamma^0 = \begin{bmatrix} ZZ^\dagger & 0 \\ 0 & Z^\dagger Z \end{bmatrix}, \quad (12.182)$$

BOX 12.9 - Properties of β matrices

One has $\beta^i = \Gamma^i \Gamma$ where Γ is the chiral gamma-matrix. Then

$$\beta^i \beta^j + \beta^j \beta^i = -2\delta^{ij},$$

i.e. β^i generate the universal Clifford algebra $\text{Cl}(9)$ [36]. Its basis elements are of the form β^I where I is an *ordered* subset of $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$

$$\beta^I \stackrel{\text{def}}{=} \beta^{i_1} \beta^{i_2} \cdots \beta^{i_k}, \quad \{i_1, i_2, \dots, i_k\} \equiv I \subset \{1, 2, 3, 4, 5, 6, 7, 8, 9\}.$$

One has [36]

$$\begin{aligned} (\beta^I)^\dagger &= (-1)^{|I|} (-1)^{|I|(|I|-1)/2} \beta^I = (-1)^{|I|(|I|+1)/2} \beta^I \\ \Rightarrow \quad &\begin{cases} \beta^I \text{ Hermitian} & \text{for } |I| = 0, 3 \pmod{4} \\ \beta^I \text{ anti-Hermitian} & \text{for } |I| = 1, 2 \pmod{4} \end{cases} \end{aligned}$$

where

$$ZZ^\dagger = \tau_0^2 + \tau_p V_p (\beta + \beta^\dagger) + \tau_p^2 V_p^2 \beta \beta^\dagger. \quad (12.183)$$

In Type IIA p is even. For p a multiple of 4, β is Hermitian, and $\beta^2 = 1$ by the same argument as in Eq. (12.73) cf. also **BOX 12.9**. The BPS bound is then

$$M \geq \tau_0 + \tau_p V_p. \quad (12.184)$$

For $p = 4k + 2$, β is anti-Hermitian, $\beta^2 = -1$, and the BPS bound is

$$M \geq (\tau_0^2 + \tau_p^2 V_p^2)^{1/2}. \quad (12.185)$$

Since the two-brane system has $\#_{\text{ND}} = p$, we have recovered the result of Sect. 12.6.1 for two-brane systems: for $p = 4k$ one 0-brane and one p -brane saturate the BPS bound and form a quarter-BPS configuration (for $k \neq 0$). For $p = 4k + 2$ they do not saturate the bound and cannot form a BPS two-object state.

We are interested in the existence of BPS bound states in the above system, i.e. *normalizable* eigenstates of M^2 which saturate the BPS bound (12.182) and carry the RR charges of one D0- and one D p -brane. We have to answer four questions:

- (i) existence of BPS bound states for a given (even) number p ;
- (ii) the number of supersymmetries preserved by the D0-D p bound state;
- (iii) the number of BPS D0-D p bound state supermultiplets;
- (iv) how the results extend to systems with the charges of m D0's and n D p 's.

D0-D0 Bound States

The BPS bound for the quantum numbers of two 0-branes is $M \geq 2\tau_0$, so a bound state must lay at the threshold of the continuous spectrum of two-particle states. The difference between a bound state of mass $2\tau_0$ and a two-particle state of the same mass is that the first one is a *normalizable* eigenstate of the mass operator: the question here is whether there exist *normalizable* states of energy $2\tau_0$, on top of the *non-normalizable* states with energy $2\tau_0$ which describe two *separate* D0 at rest.

To make the question more limpid, we compactify the 1-direction on a large circle of radius $R \gg \sqrt{\alpha'}$ and add one quantum of compact momentum

$$p_1 = 1/R. \quad (12.186)$$

In a two-body state this unit momentum is carried by one of the two D0-particles, and the minimal total energy of the system is

$$\tau_0 + \sqrt{\tau_0^2 + 1/R^2} \approx 2\tau_0 + \frac{1}{2\tau_0 R^2} + O(1/R^4). \quad (12.187)$$

Instead for a bound state of mass $2\tau_0$ the minimal energy is

$$\sqrt{4\tau_0^2 + 1/R^2} \approx 2\tau_0 + \frac{1}{4\tau_0 R^2} + O(1/R^4) \quad (12.188)$$

which is *strictly smaller* than the 2-particle energy (12.187). In the compact set-up, we have a clean separation between bound states and 2-particle states. To get a different (but physically equivalent) picture, we T -dualize the circle. The D0-brane becomes a D1-brane and the unit momentum turns into a quantum of F1 winding, so we end up with a F1/D1 system of charges (1, 2) wrapped on the dual circle of radius

$$R' = \alpha'/R. \quad (12.189)$$

We know from Sect. 12.7.1 that this string system contains half-BPS bound states forming exactly *one* ultrashort SUSY supermultiplet with 2^8 states.

More generally, we start with a system carrying n units of D0-charge. We compactify the 1-axis and boost the system to m units of compact momentum, where m is chosen to be coprime with n (e.g. we may take $m = 1$). This system is T -dual to a (m, n) -string, which contains *one* half-BPS supermultiplet of SUSY states.

To go back to our original problem we need to take $R \rightarrow \infty$. We have found *normalizable* BPS bound states with n units of D0 charge for all n and for all *finite* radius R . What we have to show is that they remain normalizable in the limit. While this is exceedingly plausible, a rigorous proof requires some work. The analysis was carried out in detail for $n = 2$ [30]. We conclude:

Fact 12.3 For all $n \in \mathbb{N}$ there exists precisely **one** $\frac{1}{2}$ -BPS supermultiplet of SUSY bound states of n D0-particles (2^8 SUSY states).

Fact 12.3: Evidence from Effective SUGRA In chap. 8, we found extremal (i.e. BPS) solutions to the e.o.m. of IIA SUGRA describing particles carrying n units of RR charges for all integer n . They represent bound states of n D0's. The SUGRA argument is robust by the usual non-renormalization theorems.

D0-D2 Bound States

The BPS bound shows that a BPS D0-D2 bound state has an energy strictly smaller than the two-brane continuum. There is some circumstantial evidence for the existence bound states: the long distance force between a D0 and a D2 is attractive, and the 0–2 string has a negative zero-point energy in the NS sector, Eq. (12.78): as shown in **Fig. 12.1a, b** this leads to a tachyon which survives the GSO projection. The tachyon signals instability toward a state with the same conserved quantum numbers and energy smaller than the two-object continuum, hence a bound state. However this argument is insufficient to establish whether the resulting state is BPS.

To circumvent the problem we play our favorite trick: we use T -duality to get a simpler description of the same physics. The D2 is wrapped on a T^2 as in Eq. (12.179). We take T -duality along the x^2 -axis: the D0-particle becomes a D-string wrapped on the B -cycle of T^2 , while the 2-brane becomes a D-string wrapped on the A -cycle. We recall that the cycles in T^2 on which the D1 wrap may be identified²⁷ with their RR charges. Thus the two-string system has total RR charge $A + B \in H_1(T^2, \mathbb{Z}) \simeq \mathbb{Z}^2$.

There is an obvious state with the same charge and lower energy: just take a *single* D-string wrapped on the cycle $A + B \in H_1(T_2, \mathbb{Z})$, see **Fig. 12.3b**.²⁸ A single D-string is a half-BPS state, and we conclude that a D0 and a D2 form *one* half-BPS supermultiplet of SUSY bound states.

Now we apply T -duality to the single D-string and go back to the original description. We know from Eq. (6.278) that the T -dual of a D1 forming an angle θ with

²⁷ Up to overall normalization.

²⁸ The physical processes producing the BPS bound state are as follows: the two D1's configuration in **Fig. 12.3a** has a tachyon mode which causes the strings to break at their crossing point and re-connect as in **Fig. 12.1a, b**. This produces a single D1 with a support homologous to $A + B$. The single D1 then evolves, changing its shape to minimize its energy, until it lays on a shorter closed curve in the homology class $A + B$, thus reaching the configuration in **Fig. 12.3b**.

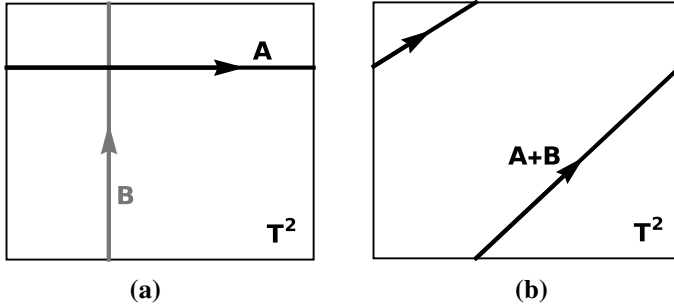


Fig. 12.3 **a** Two D1-branes wrapped, respectively, on the A and B cycles of a 2-torus T^2 drawn as a rectangle with opposite sides identified. **b** A single D1-brane wrapped on the $A + B$ cycle of T^2 . The single brane is a BPS state with the same quantum numbers as the configuration in **Fig. 12.3a**

$\tan \theta = 1$ is a D2-brane with one unit of quantized magnetic flux²⁹

$$\frac{1}{2\pi} \int_{D2} F^{(2)} = 1. \quad (12.190)$$

We check that this state has the correct quantum numbers. From the Chern–Simons action (12.50) we have

$$\begin{aligned} i\mu_2 \int_{V_{D2}} (C^{(3)} + 2\pi\alpha' F^{(2)} \wedge C^{(1)}) &\approx \\ &\approx i\mu_2 \int_{V_{D2}} C^{(3)} + i4\pi^2\alpha'\mu_2 \int_{\text{time}} C^{(1)}, \end{aligned} \quad (12.191)$$

which shows that our “D2 with flux” carries, on top of the D2-brane charge coupled to $C^{(3)}$, a D0-brane charge, *induced by the magnetic flux*, whose value is measured by its electric coupling to the RR gauge field $C^{(1)}$

$$\mu_{0, \text{induced}} = 4\pi^2\alpha'\mu_2 \equiv \mu_0, \quad (12.192)$$

to be equal to the charge μ_0 of the original D0-particle. We conclude that *the D0 has dissolved in the D2-brane in form of magnetic flux*. The mechanism which produces the BPS bound brane looks analogous to the one for the F1/D1 bound state. In fact, as we shall show in Chap. 13, while the D0-D2 and F1-D1 systems are not related by any duality visible in string perturbation theory, they do are connected by a duality at the full non-perturbative level.

²⁹ **Proof:** Take the T -dual 2-torus with radii $R'_2 = R_1$ and the D1 on the straight line $X'_2 = X_1$ which is a closed curve of minimal length in the class $A + B$. Then $A_2 \equiv X'_2/(2\pi\alpha') = X_1/(2\pi\alpha')$ so $F_{12} = 1/(2\pi\alpha')$, while $R_2 = \alpha'/R_1$. The magnetic flux is $\int_{T^2} F_{12} = 4\pi^2 R_1 R_2 F_{12} = 2\pi$.

Generalization to m D0's & n D2's We extend the analysis to bound states of m D0's and n D2's. We claim that whenever m and n are coprime integers there is a *single* ultrashort (2^8 states) BPS supermultiplet of bound states invariant under 8 supersymmetries. Indeed the same T -duality along the 2-direction produces a system of n D1's along the A -cycle and m D1's along the B -cycle, giving a total homology class $nA + mB \in H_1(T^2, \mathbb{Z})$ which is a *primitive* homology class (i.e. not a non-trivial multiple of an integral class) whenever $\gcd(m, n) = 1$. A single D1-string at an angle θ in the 1–2 plane with $\tan \theta = m/n$ gives a half-BPS configuration with the same quantum numbers iff m, n are *coprime* so that the rational number $\tan \theta$ is in minimal form. When $\gcd(m, n) = k > 1$ we get a configuration with k parallel D1's at the angle $\tan \theta = m/n$.

Note 12.7 T -dualizing back the (m, n) bound state ($\gcd(m, n) = 1$) we get a stack of n D2-branes with a $U(1)$ gauge flux $\int_{D2} F^{(2)} = 2\pi \tan \theta \equiv 2\pi m/n$. This yields a simple geometric interpretation of the quantization of the Abelian flux in multiples of $1/n$ in a $U(n)$ gauge theory which we previously deduced from 't Hooft analysis of Yang–Mills; cf. **BOX 12.8**.

D0-D4 Bound States

As with the D0-D0 case, the BPS bound (12.184) implies that any D0-D4 bound state is at the threshold, i.e. only *marginally stable*. Again the point is to distinguish normalizable and non-normalizable eigenstates of mass with the BPS eigenvalue $M = |Z|$. We can use the same trick as before, boosting the system to one quantum of compact momentum, to set a energy gap between the bound states and the multi-particle continuum.

The low-energy D0-D4 action is the 8 supercharge SQM on the D0 world-line discussed at the end of Sect. 12.6.2. Again, it is an interacting theory which becomes strongly coupled in the IR, and again the existence of SUSY bound states can be proven by non-perturbative Witten-index techniques.

Contrary to the previous cases, the BPS bound states of D0's and D4's preserve only one-quarter of the original supersymmetries: the same number as a two-object state with the same charges since only mass and charges enter in the SUSY algebra (12.180). The bound states then belong to a short supermultiplet of 2^{12} states (instead of a ultrashort 2^8 -state supermultiplet as for half-BPS states).

Claim 12.3 Let N_n be the number of BPS bound states in a system carrying the charges of n D0's and one D4 (the number of supermultiplets is then $N_n/2^{12}$). The generating function of the integral sequence³⁰ $N_0, N_1, N_2, N_3, \dots$ is

$$\sum_{n=0}^{\infty} q^n N_n = 2^8 \prod_{m=1}^{\infty} \left(\frac{1+q^m}{1-q^m} \right)^8 \equiv \frac{2^8}{\theta_4(q^2)^8} = 2^8 (1 + 16q + 144q^2 + O(q^4)). \quad (12.193)$$

³⁰ Cf. sequence A319553 in [37].

There are many ways of proving **Claim 12.3** from different viewpoints and with different standards of rigor. We may restate the **Claim**

Claim 12.4 *The generating function $Z(q)$ of the integral sequence $N_n/2^8$ is the partition function of 8 left-moving free massless scalars and 8 left-moving free massless MW fermions, i.e. the left-moving part of the R-sector partition function of the superstring (in the light cone gauge) stripped of the zero-mode factor.*

This stringy interpretation of the generating function has a deep physical meaning that we shall discuss in Chap. 13. Before going to a rigorous proof of **Claim 12.3** we sketch the physical intuition beyond it from a spacetime perspective.

Claim 12.3: Physical Heuristics We count the number of BPS bound states of n D0's and one D-brane from the spacetime viewpoint (by the previous argument, it should be a multiple of 2^{12}). This count is based on intuitive physical reasoning; a fully rigorous (and very instructive) argument will be given in the next paragraph. The weaker point in the heuristic argument is that it is not fully clear that the states we are counting are indeed normalizable eigenstates of mass.

The physical idea is that, since there is no net force between BPS objects invariant under some common set of supercharges, the several D0's do not interact between themselves nor with the D4, and for each fixed state of the D4 the system behaves as a free gas of D0-particles (as far as state counting goes). The D4 has 2^8 BPS states, while a D0-D4 two-body system has 2^{12} BPS states: therefore, for each BPS state of the D4, the single D0-particle has $2^{12}/2^8$ states. 8 of these states are bosons and 8 fermions all with a unit D0-charge. The grand-canonical partition function of a free gas of 8 bosonic and 8 fermionic states of charge 1 is

$$\mathcal{Z}(q) = \left(\frac{1+q}{1-q} \right)^8, \quad (12.194)$$

where q is the fugacity of the D0-charge. In addition we have an overall factor 2^8 from the number of BPS states of the D4-brane. This is not yet the full story, since there are other BPS configurations with the same quantum numbers. For instance, we have the two-body system of one D4 and one n D0-particle (the normalizable bound state of n D0's). The m D0-particles do not interact between themselves, nor with the D4 or other k D0-particles. So for each $m \in \mathbb{N}$ we have a free gas of 8 bosonic and 8 fermionic m D0 states, all carrying m -units of charge. The grand-canonical partition function of the free gas of m D0-particles is $\mathcal{Z}(q^m)$. Taking into account all species of m D0-particles, the full grand-canonical partition function is

$$2^8 \prod_{m \geq 1} \mathcal{Z}(q^m) = 2^8 \prod_{m=1}^{\infty} \left(\frac{1+q^m}{1-q^m} \right)^8, \quad (12.195)$$

which is **Claim 12.3**.

Exercise 12.5 Show that the coefficients N_n in the series expansion of the function (12.195) are integral multiples of 2^{12} as should be the number of $\frac{1}{4}$ -BPS states.

By T -duality the 1 D4 & n D0 system is converted into a n D4 & 1 D0 system. So the number of SUSY bound states of the new system is again N_n . More generally

Claim 12.5 *The number of BPS bound states of m D4's and n D0's is N_{mn} .*

A justification of this **Claim** will be given after the rigorous proof of **Claim 12.3**. A deeper physical proof will be given at the end of Sect. 13.2 in the next chapter.

Claim 12.3: Exact Result from the D0 World-Line Perspective We can count BPS states from the perspective of the effective theory on the D0 world-line, which is an instance of the $\#_{\text{NS}} = 4$ system studied in Sect. 12.6.2. In the present contest this becomes a 8-supercharge SQM model which is the reduction to 1d of a certain $U(n)$ gauge theory. For $n \geq 2$ it is strongly coupled at low energy. A quarter-BPS bound state of n D0-branes with one D4-brane is a *normalizable* SUSY ground state of this SQM system with one unit of D4 charge. The $\#_{\text{NS}} = 4$ system arises from the massless 0–4, 0–0, and 4–4 strings. The 4–4 d.o.f. decouple for large volume $V_4 \gg \alpha'^2$, and its 16 Fermi zero-modes generate a Clifford module with 2^8 ground states. The 0–0 d.o.f. are a $U(n)$ vector-multiplet plus a hypermultiplet Φ in the adjoint representation. The 0–4 ones give a hypermultiplet Q in the fundamental of $U(n)$. We write T_n for this SUSY $U(n)$ gauge SQM consisting of the charged hypermultiplets Q and Φ coupled to $U(n)$ SYM. The total number of BPS states is then $2^8 \cdot K_n$ where K_n is the number of normalizable, SUSY-preserving, ground states of the SQM model T_n .

It remains to compute K_n . By standard Witten-index arguments [31, 38], K_n is equal to the total dimension of the cohomology of the manifold of *classical* low-energy configurations,³¹ which, for a SQM with 8 supercharges, should be a hyper-Kähler manifold M (cf. Sect. 8.1). Since the D4-brane is compactified on a (large) torus T^4 , a classical low-energy configuration of n D0-branes is just a choice of n *undistinguishable* points on T^4 , i.e. a point in

$$(T^4)^n / \mathfrak{S}_n, \quad (12.196)$$

the symmetric n -th power of T^4 . $(T^4)^n / \mathfrak{S}_n$ is a *singular* hyperKähler space; it admits a standard resolution into a smooth *compact* hyperKähler manifold $M^{[n]}$ called its *Douady space*³², or, in the Algebro-Geometric language, its *Hilbert scheme of points* [40]. We conclude that

$$K_n = \dim H^\bullet(M^{[n]}, \mathbb{R}). \quad (12.197)$$

³¹ The statement in the text holds rigorously when the space M of classical vacua is *compact*. Then the IR description of our SQM is given by a 1d σ -model with target space the vacuum manifold M (which is Kähler for 4 or more supersymmetries) with a Hilbert space equal to the space $\Omega^\bullet(M)$ of differential forms on with their Hodge Hermitian norm $\|\psi\|^2 = \int_M \psi \wedge * \psi^*$, and Hamiltonian the Laplacian $H \equiv \Delta$; cf. the proof of **Proposition 11.3** (or [14] and references therein).

³² See Sect. 6 of [39].

There is a general formula for the dimension of the cohomology of the *resolved* symmetric n -power³³ $X^{[n]}$ of a smooth projective *surface*³⁴ X in terms of the generating function

$$\sum_{n \geq 0} q^n \dim H^\bullet(X^{[n]}) = \frac{\prod_{m=1}^{\infty} (1 + q^m)^{b_-}}{\prod_{m=1}^{\infty} (1 - q^m)^{b_+}} \quad (12.198)$$

where

$$b_+ = \dim H^{\text{even}}(X), \quad b_- = \dim H^{\text{odd}}(X), \quad (12.199)$$

see, say, Eq. (4.8) of Vafa-Witten [43] or [44, 45]. We are free to choose the lattice of $T^4 \equiv \mathbb{C}^2/\Lambda$ so that T^4 is a projective surface (an Abelian variety of dimension 2). T^4 has $b_+ = b_- = 8$; then

$$\sum_{n \geq 0} q^n K_n = \frac{\prod_{n=1}^{\infty} (1 + q^n)^8}{\prod_{n=1}^{\infty} (1 - q^n)^8}, \quad (12.200)$$

which is **Claim 12.3**. We also quote the formulae for the Euler characteristic

BOX 12.10 - More on the cohomology of the Hilbert schemes $X^{[n]}$

We write X for a smooth projective surface (over $\mathbb{C}$), $X^{(n)} \equiv X^n/S_n$ for its orbifold n -th fold symmetric product, and $X^{[n]}$ for its smooth Hilbert scheme of points, which is a smooth space. There is a birational morphism $X^{[n]} \rightarrow X^{(n)}$ which is a *crepant resolution*. We write $b_k(Y)$ for the k -th Betti number of Y , $b_k(Y) = \dim H^k(Y, \mathbb{C})$. The Poincaré polynomial of Y is

$$P(Y, z) = \sum_k b_k(Y) z^k$$

Theorem [44]. *Let X be a smooth projective surface over $\mathbb{C}$. Then*

$$\begin{aligned} \sum_{n=0}^{\infty} P(X^{[n]}, z) q^n &= \prod_{m=1}^{\infty} \frac{(1 + z^{2m-1} q^m)^{b_1(X)} (1 + z^{2m+1} q^m)^{b_1(X)}}{(1 - z^{2m-2} q^m)^{b_0(X)} (1 - z^{2m} q^m)^{b_2(X)} (1 - z^{2m+2} q^m)^{b_0(X)}} \\ &= \prod_{k=0}^4 \left(\prod_{n=1}^{\infty} \left(1 - (-z)^{2(m-1)+k} q^m \right) \right)^{(-1)^{k-1} b_k(X)}. \end{aligned}$$

Setting $z = \pm 1$ one gets the formulae in the main text. The Betti numbers of the unresolved symmetric product $X^{(n)}$ is given by restricting to the $m = 1$ factor in the above formula

³³ Which (in this case) coincides with the cohomology of the orbifold X^n/S_n in the sense of orbifold cohomology [41, 42].

³⁴ Here *surface* means a compact complex manifold of complex dimension 2. *Projective* means that it is a complex submanifold of some $\mathbb{P}^N$.

$$\sum_{n \geq 0} q^n \chi(X^{[n]}) = \frac{\prod_{n=1}^{\infty} (1 - q^n)^{b_-}}{\prod_{n=1}^{\infty} (1 - q^n)^{b_+}} = \frac{1}{\prod_{n=1}^{\infty} (1 - q^n)^{\chi(X)}}. \quad (12.201)$$

For additional details see **BOX 12.10**.

Note 12.8 The trace of the $n \times n$ matrix Φ is a free hypermultiplet, whose fermionic zero-modes produce 2^4 states. Thus K_n is a multiple of 2^4 (cf. **Exercise 12.5**).

Claim 12.5: the D4 World-Volume Perspective The description from the viewpoint of the field theory on the D4 world-volume is very deep and has several applications to diverse physical problem. It deserves a full section of its own: see Sect. 12.8.

D0-D6 Bound States

The relevant BPS bound is (12.185), and a BPS bound state, *if it exists*, would have an energy below the two-object continuum. For large separations $|y| \gg \sqrt{\alpha'}$ the force between a D0- and a D4-brane is *repulsive* (cf. **BOX 12.11**). In the opposite limit, $|y| \rightarrow 0$, the zero-point energy of 0–6 NS strings is positive (cf. Eq. (12.79))

$$E_0 = \frac{\#_{\text{ND}}}{8} - \frac{1}{2} = \frac{6}{8} - \frac{1}{2} = \frac{1}{4}, \quad (12.202)$$

and, contrary to the D0-D2 case, there is no instability which would end into a lower energy state. Hence there is no indication of a bound state from either the spacetime or the world-sheet perspective. For the D6 world-volume perspective, see *Note 12.9* below. The common conclusion from all three approaches is that there are *no* SUSY state carrying the charges of one D0 and D6.

BOX 12.11 - Sign of long-distance force between a D0 and a D2k

Claim The long-range force between a D0 and a D2k has the same sign as $-\prod_{a=1}^4 \sin \theta_a$ where

$$\begin{aligned} \theta_1 &= \frac{1}{2}(\phi_1 + \phi_2 + \phi_3 + \phi_4), & \theta_2 &= \frac{1}{2}(\phi_1 + \phi_2 - \phi_3 - \phi_4), & \phi_a &= \begin{cases} \pi/2 & a \leq k \\ 0 & a > k \end{cases} \\ \theta_3 &= \frac{1}{2}(\phi_1 - \phi_2 + \phi_3 - \phi_4), & \theta_4 &= \frac{1}{2}(\phi_1 - \phi_2 - \phi_3 + \phi_4), \end{aligned}$$

so is negative for $k = 1$, positive for $k = 3$ and zero otherwise.

Proof By a chain of T -dualities we transform the system into a pair of D4 rotated by the angles ϕ_a as in the **Example** around Eq. (12.96) By Eq. (12.118), the sign of the long-range force is the sign of $\prod_a \sin \theta_a$ times the sign of the long-range force associated with the NS-NS exchange, which is always attractive (since it is dominated by the gravitational force).

D0-D8 Bound States

D8-branes are less innocent objects than we may expect. They are subtle in a number of ways: in particular they are inherently strongly coupled. For this reason, we defer

their discussion to the end of chap. 13 once we have understood the physics of the superstring at strong coupling.

12.8 D-Branes as Yang–Mills Instantons

We return to the D0–D4 system. We consider a system of one D0 and n D4’s that we take to be on top of each other (i.e. parallel at zero separation).³⁵ Seen from the D0 world-line, this system is a 8-supercharge SQM which is the dimensional reduction to 1d of the 4d $\mathcal{N} = 2$ $U(1)$ gauge theory coupled to a charged hypermultiplet with an unbroken $U(n)$ flavour symmetry from the D4 “Chan–Paton” label. We write the hypermultiplet scalars as $\chi_{i,a}$ with $i = 1, 2$ and $a = 1, \dots, n$.

Higgs versus Coulomb Branch The scalar potential of the SQM reads

$$V = \frac{g_{\text{D0}}^2}{4} \sum_{A=1}^3 (\chi_{i,a}^\dagger \sigma_{ij}^A \chi_{j,a})^2 + \chi_{i,a}^\dagger \chi_{i,a} \sum_{s=1}^5 \frac{(X_s - X'_s)^2}{2\pi\alpha'}, \quad (12.203)$$

cf. Eqs. (12.86), (12.92). The second term by itself has two branches of zeros

$$\begin{aligned} \text{(H)} \quad & X_i - X'_i = 0, \quad \chi_{i,a} \neq 0, \\ \text{(C)} \quad & X_i - X'_i \neq 0, \quad \chi_{i,a} = 0. \end{aligned} \quad (12.204)$$

The branch **(H)**, where the hypermultiplet scalars are non-zero, is known as a *Higgs branch*. Along the Higgs branch the D4’s D0-brane is inside the D4 world-volume. The branch **(C)**, where the vector-multiplet scalars $X_i - X'_i$ are non-zero and the D0 is away from the D4’s, is known as the *Coulomb branch*. When $n = 1$ the first term in (12.203) eliminates the Higgs branch since the D-term condition

$$D^A \equiv \chi_{i,a}^\dagger \sigma_{ij}^A \chi_{j,a} = 0, \quad (12.205)$$

implies $\chi_i = 0$. When $n \geq 2$ the D-term condition does not force $\chi_{i,a}$ to vanish. For instance, for $n = 2$, the general solution of the D-term condition is

$$\chi_{ia} = v U_{ia}, \quad (12.206)$$

where v is an arbitrary complex constant and $U \in SU(2)$. We may make v real by a $U(1)$ gauge transformation. Thus for $n \geq 2$ D4-branes the space of SUSY vacua³⁶ has a Higgs branch of positive dimension. Along the **(H)** branch the D0 is on the D4’s

³⁵ We stress that this system is related to the n D0 and 1 D4 system studied non-perturbatively above by T -duality.

³⁶ For a SQM with 8 supercharges the space of *generic* classical and quantum vacua are identified.

and the R-symmetry $SU(2)_R \subset Spin(4)_{\text{ND}}$ which rotates the hypermultiplet scalars (cf. Sect. 12.6.2) is spontaneously broken, while it is unbroken in the (C) branch.

The Coulomb branch has an obvious physical/geometric interpretation: it describes the separation of the D0 from the stack of D4's in the normal direction. *What is the physical/geometric interpretation of the Higgs branch?*

The D4 World-Volume Perspective

To answer the question, we may look at the situation from the viewpoint of the field theory living on the D4's. Non-Abelian 4d gauge theories have *instantons*, that is, finite-action Euclidean-signature classical solutions localized in space-time [46–48] (for a recent review see [49]). Their field strength is self-dual or anti-self-dual (depending on the sign of the topological charge)

$$*F^{(2)} = \pm F^{(2)}, \quad (12.207)$$

so the Bianchi identity implies the field equations. Since the classical Yang–Mills theory is scale invariant, the characteristic size of the configuration is arbitrary—we have a family of classical solutions parametrized by the scale size $\rho > 0$. The $U(N)$ gauge theory on N coincident D4-branes is five-dimensional, and a configuration constant in time with $A_0 = 0$ and $A_i(x^j)$ equal to a 4d instanton configuration is a static classical solution, a soliton called the *5d instanton-particle*.

The instanton-particle has many properties in common with the D0-brane bounded to the stack of D4-branes: as we are going to show they are indeed one and the same thing. [50–52]. However at this point it may be more natural to T -dualize the system in the time direction and check the identification in the *equivalent* system of one $D(-1)$ -instanton bound to a stack of N *Euclidean signature* D3's, whose world-volume theory *per se* is 4d Euclidean $\mathcal{N} = 4$ SYM with $G = U(N)$. We write E3 instead of D3 to emphasize that the branes have Euclidean-signature world-volume. We wish to compare the system of N E3's bounded to a $D(-1)$ with a classical instanton in the 4d $\mathcal{N} = 4$ SYM theory with $G = SU(N)$.

(I) Preserved SUSY The classical YM instanton is a BPS state, breaking half of the 16 supersymmetries of $\mathcal{N} = 4$ SYM, exactly as the 1 $D(-1)$ with N D3 system. Indeed the SUSY variation of the gaugino is

$$\delta\lambda \propto F_{MN}\Gamma^{MN}\epsilon, \quad (12.208)$$

where Γ^{MN} are the generators of the $Spin(4) = SU(2) \times SU(2)$ Euclidean group. The self-duality constraint (12.207) says that only the generators of the first (or second) $SU(2)$ enter in the variation (12.208), while each of the 4 SUSY parameters ϵ^a decomposes as $(2, 1) \oplus (1, 2)$ under $SU(2) \times SU(2)$, so that $4 \times 2 \equiv 8$ SUSY remain unbroken.

(II) Quantum Numbers The instanton carries the same RR charges as the $D(-1)$. Indeed, expanding the Chern–Simons action (12.50) we get a term

$$\frac{1}{2}(2\pi\alpha')^2\mu_3 \int C^{(0)} \wedge \text{tr}\left(F^{(2)} \wedge F^{(2)}\right). \quad (12.209)$$

The topological charge of the instanton is

$$\int_{D3} \text{tr}(F^{(2)} \wedge F^{(2)}) = 8\pi^2, \quad (12.210)$$

so the total coupling to a constant $C^{(0)}$ background is

$$(4\pi^2\alpha')^2\mu_3 \equiv \mu_{-1}, \quad (12.211)$$

which is exactly the $D(-1)$ -charge; cf. Eq. (12.41).

(III) Moduli Spaces The (framed³⁷) moduli space of $SU(2)$ instantons, i.e. the space which parametrizes the self-dual solutions of $SU(2)$ YM with topological charge (12.210), matches the moduli space of BPS configurations along the branch Higgs branch $\mathbb{R}^4 \times M$, where $\mathbb{R}^4$ is the position of the $D(-1)$ in the $D3$ world-volume, parametrized by the $D(-1)$ transverse scalars X^i ($i = 1, \dots, 4$) in the directions parallel to the $D3$'s and M is the space parametrized by the scalars $\chi_{i,a}$. The scale ρ of the instanton corresponds to the overall factor v in the v.e.v. of $\chi_{i,a}$ and the instanton orientation in the gauge group corresponds to $U_{i,a} \in SU(2)$. We shall present a more precise argument for the identification of the two spaces at the end of this section.

We conclude that we must identify the $D(-1)$ on a stack of N coinciding $E3$'s with an instanton of $\mathcal{N} = 4$ SYM with gauge group $SU(n)$, and hence by T -duality a $D0$ on a stack of $D4$'s with an instanton-particle. The identification extends to m $D(-1)$'s on a stack of N Euclidean $D3$'s with a self-dual configurations in $\mathcal{N} = 4$ $SU(N)$ SYM with topological charge (instanton number) m : indeed for all m the two systems are quarter-BPS states with the same quantum numbers and also the moduli match as we show below. The identification gives a new field-theoretic proof of the existence of the $D(-1)$ as a dynamical object of Type IIB. The v.e.v. of the RR scalar is then identified with the YM instanton angle, $\theta = 2\pi \langle C^{(0)} \rangle$; cf. Eq. (12.209).

Specializing the arguments in Sect. 8.1 to 1d, we see that the space $\mathcal{M}$ of classical vacua of a SQM model, away from singularities, is a Riemannian manifold whose holonomy depends on the number $2k$ of supercharges

$2k$	2		4	8	> 8	(12.212)
$\mathcal{M}$	Riemannian		Kähler	hyperKähler	flat	

Since the instanton preserves 8 supersymmetries, as a corollary we get the following well-known and fundamental math fact

³⁷ By this we mean that we do not mod out the overall $SU(N)$ gauge transformation, i.e. we consider as distinct two configurations which differ by a rigid $SU(N)$ rotation.

Fact 12.4 The moduli space $\mathcal{M}_{k,N}$ of k instantons in $SU(N)$ gauge theory, equipped with its natural metric, is *hyperKähler*.

Below we give a math proof of this **Fact**.

Discussion The precise connection between the D0-brane and the instanton-particle is as follows [50]. When the scale size ρ is large compared to the string scale, the low-energy effective field theory on the D4-branes is a good description of the instanton. However, as ρ is reduced below the string length $\sqrt{\alpha'}$ this description breaks down. The D0-brane picture provides an alternative description which gets accurate in the opposite limit. The point $v = 0$ where the Higgs and Coulomb branch cross is the zero-size instanton: switching on the Higgs moduli expands the instanton, while turning on the Coulomb moduli moves away the D0 in the normal direction. Along the Higgs branch the D0-brane dissolves into gauge-field flux inside the D4 just as in the D0-D2 case. The identification also explains why for a single D4 there is no Higgs branch: the Abelian $U(1)$ theory has no instantons.

Note the following amazing phenomenon. Start with a large instanton, an object made out of gauge field that live on the Euclidean D3-branes. Contract it to zero size, where the vacuum branches meet, and then separate it from the D3-branes going along the Coulomb branch. The “instanton” can no longer be interpreted as being made of gauge fields, because they exist only on the D4-branes.

Back to Bound States

Returning to our original bound state problem, the system with m D0-branes bound to n D4-branes is described by SUSY quantum mechanics on the moduli space $\mathcal{M}_{m,n}$ of $SU(n)$ instantons of total topological charge m . We know that this space is hyperKähler. The number of SUSY states is related to the topology of this space, in fact it is equal to $2^8 \dim H^\bullet(\mathcal{M}_{m,n}, \mathbb{R})$ which is expected to be equal to N_{mn} see [53].

Note 12.9 We return to the existence question for D0-D6 BPS bound states, or equivalently for D(−1)-E5 ones. From the perspective of the D5 world-volume theory these configurations would look like half-BPS instantons of 6d (classical) $SU(N)$ SYM, with 16 SUSYs, which carry one unit of a topological charge proportional to $\int \text{tr}(F^{(2)})^3$. However there is no such a classical solution, since a configuration with $\int \text{tr}(F^{(2)})^3 \neq 0$ can be at most quarter-BPS.³⁸ We conclude again that no D0-D6 bound state exists.

More on the Instanton Moduli Space $\mathcal{M}_{k,N}$

The identification of instantons with D(−1) branes absorbed into E3’s allows for an explicit construction of the instantons for the classical gauge groups $SU(N)$, $SO(N)$ and $Sp(N)$ of arbitrary topological charge k which is identical to the ADHM (Atiyah–Drinfeld–Hitchin–Manin) construction [54–56] originally obtained by twistorial and

³⁸ The SUSY variations have the form $\delta\psi \propto F_{ij}\Gamma^{ij}\epsilon$, where Γ^{ij} generate $\text{Spin}(6)$. The number of unbroken SUSYs is twice the number of parallel spinor for the holonomy algebra $\mathfrak{a} \subset \mathfrak{spin}(6)$, so 4 for $\mathfrak{a} = \mathfrak{su}(3)$, 8 for $\mathfrak{a} = \mathfrak{su}(2)$, etc. But $\mathfrak{a} \subseteq \mathfrak{su}(2)$ gives $\int \text{tr}(F^{(2)})^3 = 0$.

Algebro-Geometric methods. Thus D-branes give a physical understanding of the abstract mathematical construction [57, 58]. The ADHM construction is a special case of the Nahm transform [59–61], which is a particular instance of the Mukai transform [62]. These more general constructions also have a physical interpretation [63]. We prefer to postpone the ADHM construction after the discussion of strings at strong coupling, see Sect. 14.2. Here we limit to show that the moduli space of Higgs vacua for a system of k D(−1) and N E3's, with its exact HK metric, is the same hyperKähler manifold as the instanton moduli $\mathcal{M}_{k,N}$ described by ADHM.

ADHM Data In the ADHM formalism, a $SU(N)$ instanton of topological charge k is specified by the following *ADHM data*: a $k \times N$ quaternionic matrix S which we see as a pair of $k \times N$ complex matrices Q and $\tilde{Q}$

$$S = \begin{pmatrix} Q \\ \tilde{Q} \end{pmatrix}, \quad (12.213)$$

and four Hermitian $k \times k$ matrices X_μ ($\mu = 1, 2, 3, 4$) which we write as two complex $k \times k$ matrices $B_1 = X_2 + iX_1$ and $B_2 = X_4 + iX_3$. Two ADHM data $(Q, \tilde{Q}, B_1, B_2)$ and $(Q', \tilde{Q}', B'_1, B'_2)$ produce the same instanton iff

$$Q' = UQ, \quad \tilde{Q}' = U\tilde{Q}, \quad B'_a = UB_aU^{-1}, \quad \text{for } U \in U(k), \quad a = 1, 2. \quad (12.214)$$

The ADHM data are subjected to the constraints

$$[B_1, B_2] + Q\tilde{Q}^\dagger = 0, \quad [B_1, B_1^\dagger] + [B_2, B_2^\dagger] + QQ^\dagger - \tilde{Q}\tilde{Q}^\dagger = 0. \quad (12.215)$$

The (framed) moduli space of instantons is the hyperKähler space given by the quotient (i.e. orbit space) of the data $(Q, \tilde{Q}, B_1, B_2)$, satisfying the constraints (12.215), by the $U(k)$ action (12.214). Geometrically, the resulting space $\mathcal{M}_{k,N}$ is the *hyper-Kähler quotient* of the flat hyperKähler manifold $\mathbb{H}^{kN+k^2}$ which we identify with $\mathbb{C}^{2k(N+k)}$ equipped with the three Kähler³⁹ (real symplectic) forms ω_a ($a = 1, 2, 3$)

$$\begin{aligned} \omega_1 + i\omega_2 &= i \operatorname{Tr}(dB_1 \wedge dB_2) + i \operatorname{Tr}(dQ \wedge d\tilde{Q}^\dagger), \\ \omega_3 &= i \sum_a \operatorname{tr}(dB_a \wedge dB_a^\dagger) + i \operatorname{tr}(dQ \wedge dQ^\dagger) + i \operatorname{tr}(d\tilde{Q} \wedge d\tilde{Q}^\dagger). \end{aligned} \quad (12.216)$$

The HK quotient is a nice technique to construct new HK manifolds from old ones having suitable symmetries, and has a deep relation to SUSY: for details see **BOX 12.12** and Refs. [64, 65]. The LHS of Eq. (12.215) are just the momentum maps

³⁹ As we shall show in detail in Sect. 14.1, for a HK $4n$ -fold from the knowledge of the 3 Kähler form we may reconstruct the corresponding Riemannian metric of holonomy $Sp(n)$.

$\mu_1 + i\mu_2$ and μ_3 for the $U(k)$ action (12.214) on $\mathbb{C}^{2k(N+k)}$ with respect to the three symplectic forms ω_a in (12.216). As explained in box **BOX 12.12**, the orbit space $\mathcal{M}_{k,N}$ —equipped with the metric induced from the flat HK one on $\mathbb{C}^{2k(N+k)}$ —is hyperKähler. This yields the math proof of the physical **Fact 12.4**: the (framed) instanton moduli $\mathcal{M}_{k,N}$ is actually hyperKähler of real dimension is $4Nk$.

Comparison with the k D(−1) N E3’s From the viewpoint of the D-instanton world-volume theory, our brane system is described by a zero-dimensional QFT obtained by dimensional reduction of a 4d $N = 2$ QFT with a gauge group $U(k)$ and flavour symmetry $U(N)$, coupled to an adjoint hypermultiplet, whose scalars we write as two $k \times k$ complex matrices (B_1, B_2) , and one $(\mathbf{k}, \mathbf{N})$ hypermultiplet whose bosonic d.o.f. we write as a pair of complex $k \times N$ matrices $(Q, \tilde{Q})$. The hypermultiplet

BOX 12.12 - HyperKähler quotient

Recall that a (smooth) action of a Lie group G on a symplectic manifold M , with symplectic form ω , is *Hamiltonian* iff the vector fields $v \in \mathfrak{g} \equiv \mathfrak{Lie}(G)$ leave invariant the symplectic form, i.e. $0 = \mathcal{L}_v \omega \equiv d(i_v \omega)$ (since the symplectic form is closed). Assuming (for simplicity) M simply-connected, we have $i_v \omega = d\mu(v)$ for some function $\mu(v)$, so we get a map $\mu: M \rightarrow \mathfrak{g}^\vee$ called the *momentum map* or the Hamiltonian generating the flow $\exp(vt): M \rightarrow M$. For instance, M may be the phase space of a classical Hamiltonian system invariant by $SO(2)$ rotations; the map μ is just the $SO(2)$ angular momentum as a function on the phase space. In mechanics we can separate the angular variable θ associated with the $SO(2)$ symmetry, i.e. we may reduce to a Hamiltonian system with *one less* degree of freedom which describes the “radial” motion. This is achieved by fixing the values of the momentum $\mu = J$ and then taking the orbit space of the flow $\exp(vt)$, i.e. ignoring the value of the angle θ canonically dual to μ . This procedure works for all Hamiltonian actions on symplectic manifolds (under some mild regularity condition). Mathematicians call it the *symplectic quotient* (or the Marsden–Weinstein quotient [66, 67]) and produces a symplectic manifold (“phase space”) of dimension $\dim M - 2 \dim G$.

Let now M be a HK manifold with its 3 Kähler form ω_a which are 3 symplectic forms transforming in the adjoint of $\mathfrak{sp}(1)$. The action of a Lie group G on M is called *hyperHamiltonian* iff it is Hamiltonian for all 3 symplectic forms, i.e. iff $\mathcal{L}_v \omega_a = 0$ for $a = 1, 2, 3$ and $v \in \mathfrak{g}$. In this case we get a triplet of momentum maps $\mu_a(v)$ such that $i_v \omega_a = d\mu_a(v)$, or in another notation, a *hyperKähler momentum map*

$$\mu: M \rightarrow \mathfrak{g}^\vee \otimes \mathfrak{sp}(1).$$

The Hamiltonian separation of variables ($\equiv$ Marsden–Weinstein theorem) has a HK analogue.

Theorem [64, 65] *M a HK with a hyperHamiltonian action by G and HK momentum map μ . Let $\lambda \in \mathfrak{g}^\vee \otimes \mathfrak{sp}(1)$ be fixed by the co-adjoint action of G , and suppose that $N(\lambda) = \mu^{-1}(\lambda) \subset M$ is a manifold and the orbit space $M(\lambda) \equiv N/G$ is a manifold (orbifold). Then $M(\lambda)$ is a HK manifold (orbifold) whose HK structure is induced from M via inclusion and projection maps.*

The HK space $M(\lambda)$ is called the *hyperKähler quotient* of M by G (at $\mu = \lambda$). One has $\dim_R M(\lambda) = \dim_{\mathbb{R}} M - 4 \dim G$. HK quotient is a handy technique to construct new HK manifold from given ones with symmetry.

(B_1, B_2) , instead, describe motions of the $D(-1)$ parallel⁴⁰ to the $E3$'s. The 0d theory contains, in addition, other six scalars X^i ($i = 0, 4, \dots, 9$), coming from the 4d $\mathcal{N} = 2$ vector-multiplet, which describe the motion of the $D(-1)$ branes *away* from the $E3$'s. These vector-multiplet scalars are zero along the Higgs branch (they parametrize the Coulomb branch). The bosonic part of the 0d action is just a scalar potential V . The potential which correspond to our brane system can be obtained by dimensional reduction of the 4d one. Setting to zero the vector-multiplet scalars, V is (up to overall normalization)

$$V \propto \text{Tr}([B_1, B_1^\dagger] + [B_2, B_2^\dagger] + QQ^\dagger - \tilde{Q}\tilde{Q}^\dagger)^2 + \\ + \text{Tr}([B_1, B_2] + Q\tilde{Q}^\dagger)([B_1, B_2] + Q\tilde{Q}^\dagger)^\dagger]. \quad (12.217)$$

In the 4d $\mathcal{N} = 1$ language the first term is the square of the D-term and the second one the square of the F-term. The zero-dimensional vacua along the Higgs branch are just the zeros of this non-negative potential, which are precisely the solutions to Eq. (12.216). Thus the Higgs branch of the 0d QFT is a hyperKähler manifold which precisely coincides with the ADHM moduli of $SU(N)$ instantons (before modding out the overall $SU(N)$) with the same natural metric of holonomy $Sp(kN)$.

More details on the ADHM construction and D-branes in Sect. 14.2.

References

1. Deser, S., Zumino, B.: Consistent supergravity. *Phys. Lett. B* **62**, 335 (1976)
2. Coleman, S.R.: There are no Goldstone bosons in two-dimensions. *Commun. Math. Phys.* **31**, 259–264 (1973)
3. Haag, R., Łopuszanski, J.T., Sohnius, M.: All possible generators of supersymmetries of the S -matrix. *Nucl. Phys. B* **88**, 257 (1975)
4. Strathdee, J.: Extended Poincaré supersymmetry. *Int. J. Mod. Phys. A* **2**, 273–300 (1987)
5. Coleman, S.R.: The magnetic monopole fifty years later. In: *Proceedings Les Houches Summer School in Theoretical Physics. Gauge Theories in High Energy Physics. Les Houches, vol. 1982*, pp. 461–552 (1981)
6. Witten, E.: Toroidal compactification without vector structure. *JHEP* **02**, 006 (1998). [arXiv:hep-th/9712028](https://arxiv.org/abs/hep-th/9712028)
7. Dabholkar, A.: Lectures on Orientifolds and Duality. [arXiv:hep-th/9804208](https://arxiv.org/abs/hep-th/9804208)
8. Distler, J., Freed, D.S., Moore, G.W.: Orientifold précis. In: *Mathematical Foundations of Quantum Field Theory and Perturbative String Theory*, vol. 83 of *Proceedings of Symposia in Pure Mathematics*, pp. 159–172. American Mathematical Society, Providence, RI (2011). [arXiv:0906.0795](https://arxiv.org/abs/0906.0795)
9. Gao, D., Hori, K.: On the structure of the Chan-Paton factors for D-Branes in Type II orientifolds. [arXiv:1004.3972](https://arxiv.org/abs/1004.3972)
10. Polchinski, J.: Dirichlet branes and Ramond-Ramond charges. *Phys. Rev. Lett.* **75**, 4724–4727 (1995). [arXiv:hep-th/9510017](https://arxiv.org/abs/hep-th/9510017)

⁴⁰ The $k \times k$ matrices X^μ associated with the parallel motions of the k $D(-1)$ are $X^2 + iX^i \equiv B_1$ and $X^4 + iX^3 \equiv B_2$.

11. Polchinski, J.: TASI lectures on D-branes. In: Efthimiou, C., Greene, B. (eds.) *Fields, Strings, and Duality*, TASI 1996, pp. 293–356. World-Scientific (1997). [arXiv:hep-th/9611050](#)
12. Dirac, P.A.M.: Quantised singularities in the electromagnetic field. *Proc. Roy. Soc. Lond. A* **133**, 60–72 (1931)
13. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](#)
14. Cecotti, S.: *Supersymmetric Field Theories. Geometric Structures and Dualities*. Cambridge University Press (2015)
15. Tseytlin, A.A.: Born-Infeld action, supersymmetry and string theory. [arXiv:hep-th/9908105](#)
16. Gibbons, G.W.: Born-Infeld particles and Dirichlet p-branes. *Nucl. Phys. B* **514**, 603–639 (1998). [arXiv:hep-th/9709027](#)
17. Callan, C.G., Maldacena, J.M.: Brane death and dynamics from the Born-Infeld action. *Nucl. Phys. B* **513**, 198–212 (1998). [arXiv:hep-th/9708147](#)
18. Howe, P.S., Lambert, N.D., West, P.C.: The selfdual string soliton. *Nucl. Phys. B* **515**, 203–216 (1998). [arXiv:hep-th/9709014](#)
19. Lee, S., Peet, A.W., Thorlacius, L.: Brane waves and strings. *Nucl. Phys. B* **514**, 161–176 (1998). [arXiv:hep-th/9710097](#)
20. Emparan, R.: Born-Infeld strings tunneling to D-branes. *Phys. Lett. B* **423**, 71–78 (1998). [arXiv:hep-th/9711106](#)
21. Gauntlett, J.P., Gomis, J., Townsend, P.K.: BPS bounds for world volume branes. *JHEP* **01**, 003 (1998). [arXiv:hep-th/9711205](#)
22. Amati, D., Ciafaloni, M., Veneziano, G.: Can space-time be probed below the string size? *Phys. Lett. B* **216**, 41 (1989)
23. Fabbrichesi, M., Veneziano, G.: Thinning out of relevant degrees of freedom in scattering of strings. *Phys. Lett. B* **233**, 135 (1989)
24. Konishi, K., Paffuti, G., Provero, P.: Minimum physical length and the generalized uncertainty principle in string theory. *Phys. Lett. B* **234**, 276 (1990)
25. Yoneya, T.: Duality and indeterminacy principle in string theory. In: Kawai, K., Ueda, A. (eds.) *Wondering in the Fields*, p. 419. World Scientific (1987)
26. Yoneya, T.: On the interpretation of minimal length in string theories. *Mod. Phys. Lett. A* **4**, 1587 (1989)
27. Polchinski, J.: *String Theory*, vols. 1, 2. Cambridge University Press, Cambridge (2003)
28. Berkooz, M., Douglas, M.R., Leigh, R.G.: Branes intersecting at angles. *Nucl. Phys. B* **480**, 265 (1996). [arXiv: hep-th/9606139](#)
29. Witten, E.: Bound states of strings and p-branes. *Nucl. Phys. B* **460**, 335–350 (1996). [arXiv:hep-th/9510135](#)
30. Sethi, S., Stern, M.: D-brane bound states redux. *Commun. Math. Phys.* **194**, 675–705 (1998). [arXiv:hep-th/9705046](#)
31. Witten, E.: Constraints on supersymmetry breaking. *Nucl. Phys. B* **202**, 253–316 (1982)
32. Hooft, G.: A property of electric and magnetic flux in non-Abelian gauge theories. *Nucl. Phys. B* **153**, 141–160 (1979)
33. Hooft, G.: Confinement and topology in non-Abelian gauge theories. *Acta Phys. Austriaca Suppl.* **22**, 531–586 (1980)
34. Witten, E.: θ vacua in two-dimensional quantum chromodynamics. *Nuovo Cim. A* **51**, 325 (1979)
35. Hooft, G.: Naturalness, chiral symmetry, and spontaneous chiral symmetry breaking. *NATO Sci. Ser. B* **59**, 135–157 (1980)
36. Postnikov, M.M.: *Lectures in Geometry: Lie Groups and Lie Algebras*. Editorial URSS (1994)
37. On-line Encyclopaedia of Integer Sequence. <https://oeis.org>
38. Witten, E.: Supersymmetry and Morse theory. *J. Diff. Geom.* **17**(4), 661–692 (1982)
39. Beauville, A.: Variétés Kähleriennes dont la première classe de Chern est nulle. *J. Diff. Geom.* **18**, 755–782 (1983)
40. Göttsche, L.: *Hilbert schemes of zero-dimensional subschemes of smooth varieties*. Lecture Notes in Mathematics, vol. 1572. Springer, Berlin (1994)

41. Dixon, L.J., Harvey, J.A., Vafa, C., Witten, E.: Strings on orbifolds. Nucl. Phys. B **261**, 678–686 (1985)
42. Dixon, L.J., Harvey, J.A., Vafa, C., Witten, E.: Strings on orbifolds (II). Nucl. Phys. B **274**, 285–314 (1986)
43. Vafa, C., Witten, E.: A strong coupling test of S -duality. Nucl. Phys. B **431**, 3–77 (1994)
44. Göttsche, L.: The Betti numbers of the Hilbert scheme of points on a smooth projective surface. Math. Ann. **286**, 193–207 (1990)
45. Hirzebruch, F., Höfer, T.: On the Euler number of an orbifold. Math. Ann. **286**, 255–260 (1990)
46. Belavin, A.A., Polyakov, A.M., Schwartz, A.S., Tyupkin, Y.S.: Pseudoparticle solutions of the Yang-Mills equations. Phys. Lett. B **59**, 85–87 (1975)
47. Hooft, G.: Computation of the quantum effects due to a four-dimensional pseudoparticle. Phys. Rev. **D14**, 3432–3450 (1976)
48. Coleman, S.R.: The uses of instantons. Subnucl. Ser. **15**, 805 (1979)
49. Vandoren, S., van Nieuwenhuizen, P.: Lectures on instantons. [arXiv:0802.1862](https://arxiv.org/abs/0802.1862)
50. Witten, E.: Small instantons in string theory. Nucl. Phys. B **460**, 541–559 (1996). [arXiv:hep-th/9511030](https://arxiv.org/abs/hep-th/9511030)
51. Douglas, M.R.: Branes within branes. NATO Sci. Ser. C **520**, 267–275 (1999). [arXiv:hep-th/9512077](https://arxiv.org/abs/hep-th/9512077)
52. Ganor, O.J., Hanany, A.: Small $E(8)$ instantons and tensionless noncritical strings. Nucl. Phys. B **474**, 122–140 (1996). [arXiv:hep-th/9602120](https://arxiv.org/abs/hep-th/9602120)
53. Vafa, C.: Instantons on D-branes. Nucl. Phys. B **463**, 435 (1996). [arXiv:hep-th/9512078](https://arxiv.org/abs/hep-th/9512078)
54. Atiyah, M.F., Drinfeld, V.G., Hitchin, N.J., Manin, Y.I.: Phys. Lett. A **65**, 185 (1978)
55. Atiyah, M.F.: Geometry of Yang-Mills Fields. Lezioni Fermiane, Scuola Normale Superiore (1979)
56. Corrigan, E., Goddard, P.: Construction of instanton and monopole solutions and reciprocity. Ann. Phys. **154**, 253 (1984)
57. Witten, E.: Sigma models and the ADHM construction of instantons. J. Geom. Phys. **15**, 215–226 (1995)
58. Douglas, M.R.: Gauge fields and D-branes. J. Geom. Phys. **28**, 255 (1998). [arXiv:hep-th/9604198](https://arxiv.org/abs/hep-th/9604198)
59. W. Nahm, Self-dual monopoles and calorons, in Group theoretical methods in physics (Trieste: Lecture Notes in Physics 201, vol. 1984, pp. 189–200. Springer, Berlin (1983)
60. Schenk, H.: On a generalised Fourier transform of instantons over flat tori. Comm. Math. Phys. **116**, 177–183 (1988)
61. Braam, P.J., van Baal, P.: Nahm’s transformation for instantons. Comm. Math. Phys. **122**, 267–280 (1989)
62. Bartocci, C., Bruzzo, U., Ruipérez, D.H.: Fourier-Mukai and Nahm transforms in geometry and mathematical physics. Birkhäuser (2009)
63. Cecotti, S., Gaiotto, D., Vafa, C.: tt^* geometry in 3 and 4 dimensions. JHEP **05**, 055 (2014). [arXiv:1312.1008](https://arxiv.org/abs/1312.1008)
64. Hitchin, N.J., Krlhede, A., Lindström, U., Roček, M.: Hyper-Kähler metrics and supersymmetry. Comm. Math. Phys. **108**, 535–589 (1987)
65. Boyer, C., Galicki, K.: Sasakian Geometry. Oxford Mathematical Monographs. Oxford University Press (2008)
66. Marsden, J., Weinstein, A.: Reduction of symplectic manifolds with symmetry. Rep. Mathematical Phys. **5**, 121–130 (1974)
67. Cannas da Silva, A.: Lectures on Symplectic Geometry. Lectures Notes in Mathematics, vol. 1764. Springer, Berlin (2008)

Chapter 13

SUSY Strings at Strong Coupling



Abstract We study the five 10d SUSY string theories at strong coupling. In this regime each model has a dual description either by a weakly coupled string theory or by a new quantum system, living in 11 dimensions, called M-theory. The five 10d superstrings and M-theory are connected by a rich web of dualities. 10d IIB has a $SL(2, \mathbb{Z})$ group of auto-dualities mapping strong to weak coupling. Strongly coupled IIB is dual to weakly coupled IIB with the fundamental string replaced by the D-string. The action of $SL(2, \mathbb{Z})$ on the IIB extended BPS objects is discussed. Under toroidal compactification the $SL(2, \mathbb{Z})$ duality combines with T -duality to produce a large group of U -dualities with interesting geometric, arithmetic, and physical implications. Strongly coupled Type I is dual to weakly coupled heterotic $SO(32)$ and vice versa. Type IIA at strong coupling is dual to M-theory whose low energy theory is 11d SUGRA. We discuss the BPS objects of M-theory and their relation with the BPS objects of IIA. Heterotic $E_8 \times E_8$ is dual at strong coupling to M-theory on a segment between two walls.

Thus far we have understood the string dynamics only in perturbation theory. Collective and non-perturbative phenomena, such as quark confinement and dynamical symmetry breaking, are crucial in QFT. String theory contains QFT as a subsector, and all the usual field-theoretic non-perturbative phenomena should take place in string theory as well. In addition, we expect new ‘stringy’ non-perturbative effects which we need to understand before connecting string theory to real world physics. At a conceptual level, we have a more serious issue; the perturbative series is at best asymptotic, and does not define the theory: we *need* a non-perturbative formulation.

To understand string theory beyond weak coupling, we use non-perturbative dualities. We close these introductory remarks with an *informal* discussion of them.

Dualities in Quantum Gravity Suppose we are given a low-energy effective model which contains Einstein gravity in $d \geq 4$ dimensions. Almost all such models do *not* admit a UV completion into a quantum consistent theory of gravity, i.e. they belong to the *swampland* [1]. The subset of effective theories which do have a completion is extremely sparse, possibly finite. Nevertheless let us assume that our given effective theory belongs to this “magical” subset and *does have* a consistent non-perturbative completion. It is natural to ask whether this completion is *unique*. Given that the

existence of *one* consistent completion is already a highly overdetermined condition, with only a tiny set of “magical” solutions, the existence of more than one non-perturbative completion for the same effective theory seems an impossibly hard condition. One is led to the

Tentative Physical Principle: Almost all low-energy effective theories, containing Einstein gravity in $d \geq 4$ dimensions, **do not have** a quantum-consistent non-perturbative completion. When a completion exists, it is “**unique**”

This **Tentative Physical Principle** has some dramatic consequences:

(I) It says that quantum gravity is a *predictive* physical theory in the sense [2] that to predict (in principle!) the outcome of *all* experiments one needs to perform only *finitely many* measurements, namely the ones required to fix the low-energy Lagrangian $\mathcal{L}_{\text{eff}}$. Uniqueness then determines the full infinite list of higher couplings;

(II) Two perturbatively consistent string theories which lead to the same low-energy theory should have the same non-perturbative completion, hence be the *same* quantum system. The argument works nicely when the model has more than 8 supersymmetries¹: in this case the low-energy effective Lagrangian is *exactly* known by virtue of the non-renormalization theorems (cf. Chap. 8), and we can say with certainty whether the effective theories of two models are identical or not. For instance, in Chap. 8 we saw that the $SO(32)$ heterotic and the Type I superstrings have the same low-energy description, hence non-perturbatively they should be the same theory. Nevertheless we saw in Chap. 7 that they are quite different in perturbation theory. How we reconcile this perturbative fact with our **Tentative Physical Principle**?

The *exact* low-energy theory of these two string models has a one-parameter family of 10d Poincaré invariant vacua preserving 16 SUSYs which are parametrized by the v.e.v. of the scalar Φ of 10d (1, 0) SUGRA. As $\Phi \rightarrow \pm\infty$ the tension of a 1-brane SUGRA solution goes to zero when measured in Planck units; equivalently the Planck mass goes to infinity in units of the light 1-brane tension. At the energy scale of the light 1-brane all supergravity and gauge interactions get parametrically small as $\Phi \rightarrow \pm\infty$, and the states of the light 1-brane have a sound perturbative description. Thus the unique non-perturbative theory has *two* distinct perturbative limits, $\Phi \rightarrow +\infty$ and $\Phi \rightarrow -\infty$, with different physics: the light 1-brane is the heterotic string in the first limit and the Type I string in the second one, see Sect. 13.4 for details. This is an instance of an universal pattern of *emergency of strings* at infinity in effective field space [3, 4]. The relation between two weakly coupled descriptions, in terms of different light d.o.f., of a single non-perturbative theory is dubbed a *duality*. Compare with the dual descriptions in terms of light momentum and winding modes under T -duality, Chap. 12. The **Tentative Physical Principle** plus the SUSY non-renormalization theorems imply a rich web of dualities between the several perturbative SUSY string theories. These dualities determine the strong coupling

¹ The case of 8 SUSYs was discussed in Chap. 11.

limit of the several string theories, which get unified in a *single* non-perturbative theory. The most extraordinary result is that the non-perturbative theory has a limit where *space-time is eleven-dimensional*.²

(III) Suppose we have a low-energy theory containing Einstein gravity in $d \geq 3$ which admits a consistent non-perturbative completion. All its exact symmetry (making sense in *all* topological sectors³) uplift to a symmetry of the full non-perturbative completion.⁴ The uplift is necessarily a *gauged* symmetry.

Non-perturbative dualities were introduced in [5, 6]. For reviews: [7–9].

13.1 Type IIB Strings at Strong Coupling: $SL(2, \mathbb{Z})$ Duality

In Type IIB we have two kinds of “basic” strings: the fundamental string F1, with tension $\tau_{(1,0)} \equiv 1/(2\pi\alpha')$, and the D-string D1 with tension $\tau_{(0,1)} \equiv \tau_{(1,0)}/g$. We know from Sect. 12.7.1 that, for all pair of coprime integers (p, q) , we have a half-BPS (p, q) -string carrying p units of F1 charge and q units of D1 charge. We are interested in the light degrees of freedom which propagate on the world-sheet of a (p, q) -string.

Fact 13.1 The (physical) light d.o.f. propagating on the world-sheet of a (p, q) -string are the **same** ones for all coprime integers (p, q) : transverse scalars X^i in the $\mathbf{8}_v$ of $SO(8)$, a left-moving fermion in the $\mathbf{8}_s$, and a right-moving one in the $\mathbf{8}_c$.

We give *three* proofs of Fact 13.1.

Proof (I) Consider a straight infinite (p, q) -string parallel to the 1-axis. This configuration breaks the Lorentz group $SO(9, 1) \rightarrow SO(1, 1) \times SO(8)$, breaks translational invariance in the 8 transverse directions and, being half-BPS, breaks 16 SUSYs. The light degrees of freedom are the goldstone and goldstino fields of the broken symmetries which form representations of the unbroken $SO(1, 1) \times SO(8)$. The $\mathbf{16}_s$ of $SO(9, 1)$ decomposes under the unbroken subgroup as

$$\mathbf{16}_s \rightarrow \left(\frac{1}{2}, \mathbf{8}_s\right) \oplus \left(-\frac{1}{2}, \mathbf{8}_c\right), \quad (13.1)$$

so that the left-movers (right-movers) transform in the $\mathbf{8}_s$ (resp. $\mathbf{8}_c$) □

² Recall that 11 is the largest dimension where we may have (simultaneously): (i) supersymmetry, (ii) a space-time of Lorentzian signature $(d-1, 1)$, and (iii) a propagating massless graviton.

³ Here it is absolutely crucial that the definition of the low-energy theory contains a specification of the allowed topological sectors. In particular, a symmetry of the low-energy equations of motion is a symmetry of the low-energy theory iff it maps an allowed sector into an allowed one.

⁴ More precisely, the statement is expected to hold modulo *finite* groups.

Proof (II) The light modes propagating on a (p, q) -string world-sheet can be obtained by expanding the IIB SUGRA e.o.m. around the SUGRA solution which describes the fields sourced by a (p, q) -string (cf. Sect. 8.8). Since the IIB SUGRA e.o.m. are invariant under a $SL(2, \mathbb{Z})$ symmetry, which acts transitively on the pair (p, q) of coprime integers, the expansion in light modes around the (p, q) -string for the different pair (p, q) are related by the symmetry, hence equal. $\square$

Proof (III) We determine directly the massless excitation moving on the world-sheet of an infinite D1. These excitations come from the massless states of the open string ending on the D1. The gauge field has no dynamics in 2d, and the only propagating bosonic excitations are the transverse fluctuations X^i . The Dirac equation for the massless R sector states (cf. Eqs. (3.81)–(3.86))

$$(\Gamma^0 \partial_0 + \Gamma^1 \partial_1)u = 0 \quad (13.2)$$

implies

$$\Gamma^0 \Gamma^1 u = \pm u, \quad \begin{cases} + & \text{left-movers} \\ - & \text{right-movers} \end{cases} \quad (13.3)$$

so the 2d boost operator $\frac{1}{2}\Gamma^0\Gamma^1$ has eigenvalue $s_0 = \pm\frac{1}{2}$. The open string R sector ground states transform in the $\mathbf{16}_s$ of $SO(9, 1)$ which decomposes as in Eq. (13.1) so the left-moving fermionic open string on the D-string is in an $\mathbf{8}_s$ of $SO(8)$ and the right-movers in a $\mathbf{8}_c$.

For the F1 the massless bosonic fluctuations are again the transverse fluctuations. The fermionic ones are the goldstinos. The SUSY algebra for a long string is given in Eq. (12.139) with $(p, q) = (1, 0)$

$$\frac{1}{2} \left\{ \begin{bmatrix} Q_\alpha \\ \bar{Q}_\alpha \end{bmatrix}, \begin{bmatrix} Q_\beta^\dagger \\ \bar{Q}_\beta^\dagger \end{bmatrix} \right\} = M_{\alpha\beta} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{L_1}{2\pi\alpha'} (\Gamma^0 \Gamma^1)_{\alpha\beta} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (13.4)$$

The broken SUSYs are those whose anticommutators do not vanish when acting on the BPS states; for the IIB F-string these are the Q_α with $\Gamma^0\Gamma^1 = +1$ and the $\bar{Q}_\alpha$ with $\Gamma^0\Gamma^1 = -1$. The decomposition (13.1) then shows that the Goldstone fermions on the IIB F1 have the quantum numbers in Fact 13.1. $\square$

S-Duality

We first consider strings in a trivial RR background with $C^{(0)} = 0$. The D1 and F1 strings carry the same massless d.o.f. but their tensions are different

$$\frac{\tau_{\text{F1}}}{\tau_{\text{D1}}} = g = e^\Phi. \quad (13.5)$$

At weak coupling the F1 is much lighter than the D1. The relation (13.5) follows from supersymmetry, so it is *exact* as stressed in Chap. 8. The SUSY non-renormalization arguments of Sect. 8.2 imply that, even at very strong coupling $g \gg 1$, the D1 string is a half-BPS physical object of Type IIB. But now the D1 is *much lighter* than the F1. This suggests that the IIB theory at strong coupling is equivalent to weakly-coupled IIB at coupling $1/g \ll 1$ where the two strings have reversed their role: $\text{F1} \leftrightarrow \text{D1}$.

Let us look more seriously at this idea. Define the gravitational length as

$$l_{\text{Pl}} = (4\pi^3)^{-1/8} \kappa^{1/4} \quad (\text{Planck length}), \quad (13.6)$$

where the $O(1)$ numerical constants are chosen for later convenience. The three basic length scales are in the ratios (cf. Box 13.1)

$$\tau_{F1}^{-1/2} : l_{P1} : \tau_{D1}^{-1/2} = g^{-1/4} : 1 : g^{1/4}. \quad (13.7)$$

BOX 13.1 Couplings in Type IIB

We have (Chap. 12)

$$\tau_{F1} = \frac{1}{2\pi\alpha'}, \quad \tau_{D1} = \frac{1}{2\pi\alpha'g}, \quad \kappa^2 = \frac{1}{2}(2\pi)^7 g^2 \alpha'^4.$$

Then

$$\tau_{F1}^{-1/2} = (2\pi\alpha')^{1/2}, \quad \tau_{D1}^{-1/2} = (2\pi\alpha')^{1/2} g^{1/2}, \quad l_{P1} = (2\pi\alpha')^{1/2} g^{1/4}.$$

We start at long distance and scan the physics at increasingly shorter scales. Before reaching the scale l_{P1} where gravity becomes strong, at weak coupling ($g \ll 1$) we encounter the F1 length $\tau_{F1}^{-1/2}$ which sets the mass scale of the states of the fundamental string. At this scale the gravity interactions are weak, and perturbation theory get asymptotically reliable for the F1 states. At strong coupling, $g \gg 1$, we still encounter a fundamental length before reaching the scale where gravity is strong, namely the D1 length $\tau_{D1}^{-1/2}$ which sets the mass scale of the states of the D1. Gravity is again negligible at this scale, so the D1 interactions are weak, and a perturbative treatment of the D1 string state should make sense. The new perturbative physics looks much the same as at weak coupling: gravity is weak at the new string scale, and all BPS states (which are exactly known) can be easily identified with the ones at weak coupling. At low energy the interactions are weak, and we can easily check that the identification extends also to non-BPS states. The natural conclusion is

Type IIB has a symmetry (*S-duality*) which flips $g \leftrightarrow 1/g$ and $F1 \leftrightarrow D1$

S-duality maps strong to weak coupling and vice versa. Below we shall provide further evidence for *S*-duality and test it in a variety of ways.

Note 13.1 The D-string has many massive string excitations. They play no role in the above argument, since they are not supersymmetric and decay at a rate of order g^2 . As $g \rightarrow \infty$ they become states containing a huge number of massless particles.

A priori one would think that at strong coupling we have a phase where gravity is strongly coupled and spacetime has an exotic quantum structure without a simple geometric interpretation. Instead, the strong coupling physics is identical to the weak

coupling one, up to a $SL(2, \mathbb{Z})$ rotation of the degrees of freedom. In between, for $g \approx 1$, we have no weakly coupled description, and we have no detailed control of the dynamics except in the BPS sector.

S-Duality & Low-Energy SUGRA We already know from proof (II) of Fact 13.1 that the low-energy IIB action, exact to the full non-linear level, is invariant under S -duality. $g \equiv \exp(\Phi)$, so $g \leftrightarrow 1/g$ is a flip in the sign of the dilaton

$$\Phi \longleftrightarrow -\Phi. \quad (13.8)$$

In a background with vanishing RR scalar $C^{(0)}$, the IIB e.o.m. are invariant under

$$\Phi' = -\Phi \quad G'_{\mu\nu} = e^{-\Phi} G_{\mu\nu} \quad (13.9)$$

$$B^{(2)'} = C^{(2)} \quad C^{(2)'} = -B^{(2)} \quad C^{(4)'} = C^{(4)}. \quad (13.10)$$

The Einstein frame metric

$$G_{E\mu\nu} = e^{-\Phi/2} G_{\mu\nu} = e^{-\Phi'/2} G'_{\mu\nu}, \quad (13.11)$$

is fixed by the involution (13.9). Hence the spacetime geometry, as probed by graviton wave-packets, is equal in the original and S -dual descriptions (cf. Eq. (13.7)).

13.1.1 $SL(2, \mathbb{Z})$ Duality

To get the physics of IIB as $g \rightarrow \infty$ we focused on two special BPS strings: the $(1, 0)$ - and $(0, 1)$ - ones. This is quite undemocratic: we should treat all (p, q) -strings on the same footing [10, 11]. From Fact 13.1 we know that the same light d.o.f. propagate on their world-sheet, and anyone of them may play the role of the “fundamental” string in the asymptotic limit where its tension $\tau_{(p,q)}$ goes to zero in units of the gravitational scale l_{Pl}^{-2} . In a purely NS-NS background the lightest string is either F1 or D1. However, if we switch on a suitable RR scalar $C^{(0)}$, any (p, q) -string may become the lightest string. As in Sect. 8.5 we work with the complex scalar field

$$\tau = C^{(0)} + i e^{-\Phi} \quad \text{i.e.} \quad \tau_1 = C^{(0)}, \quad \tau_2 = e^{-\Phi}, \quad (13.12)$$

taking values in the upper half-plane $\mathcal{H}$. The tension of the (p, q) -string, measured in Planck units l_{Pl}^{-2} , is given by the Hodge norm (Box 6.5)

$$\tau_{(p,q)}^2 = \|(p, q)\|_{\text{Hodge}}^2 l_{\text{Pl}}^{-4} = \frac{|p - \tau q|^2}{\tau_2 l_{\text{Pl}}^4} \equiv \frac{|p - \tau q|^2}{(2\pi\alpha')^2}. \quad (13.13)$$

The (p, q) -string gets light as we approach the *cusp* at

$$\tau = \frac{p}{q} \in \mathbb{P}^1(\mathbb{Q}) \equiv \partial \overline{\mathcal{H}} \quad (13.14)$$

in the Satake boundary [12] of the (compactified) upper half-plane $\overline{\mathcal{H}} = \mathcal{H} \cup \mathbb{P}^1(\mathbb{Q})$. In the vicinity of this cusp, we have a good string perturbation series where the role of the weakly coupled “fundamental” string is played by the (p, q) -string. The symmetry group $SL(2, \mathbb{Z})$ permutes transitively the cusps in $\mathbb{P}^1(\mathbb{Q})$ and maps one description into the other. In particular, the physics at all cusps is the same. We refer to this non-perturbative action of the modular group as the $SL(2, \mathbb{Z})$ *duality*. The S -duality (13.9)–(13.10) is just the special case of the element

$$S \equiv \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in SL(2, \mathbb{Z}) \quad (13.15)$$

restricted to its fixed line $\text{Re } \tau \equiv C^{(0)} = 0$. We already know that the $SL(2, \mathbb{Z})$ duality is consistent with the low energy SUGRA which also has a $SL(2, \mathbb{Z})$ duality with the same action on the light fields. In particular, $SL(2, \mathbb{Z})$ acts transitively on the solutions to the IIB SUGRA e.o.m. which represent half-BPS (p, q) -string.

The $SL(2, \mathbb{Z})$ duality *predicts* the existence of (p, q) -string. More precisely it predicts the existence of precisely one half-BPS supermultiplet of string states carrying the string charges (p, q) for all coprime p, q . We know from Sect. 12.7.1 that this prediction is correct: the deep non-perturbative analysis of that section may be regarded as a careful check of the duality from the perspective of the QFT living on the world-sheet.

Parabolic Subgroups Since the physics is the same at all cusps, we may focus on the cusp at infinity $\tau = i\infty$, that is, on the standard weak-coupling limit $g = C^{(0)} = 0$. The parabolic subgroup $P_{i\infty} \subset SL(2, \mathbb{Z})$ fixing the cusp at infinity,

$$P_{i\infty} = \pm \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad b \in \mathbb{Z} \quad (13.16)$$

leaves the dilaton invariant, hence it maps weak coupling to weak coupling, while shifting $C^{(0)}$ by an integer

$$C^{(0)} \rightarrow C^{(0)} + b. \quad (13.17)$$

This shift must be a symmetry already in perturbation theory, since it maps ℓ -loop contributions into ℓ -loop ones. This prediction agrees with our findings in Chap. 3: the RR scalar $C^{(0)}$ enters in the vertices only through its field strength (gradient) $F^{(1)} = dC^{(0)}$, so a shift of $C^{(0)}$ by a real constant is a symmetry of perturbation theory. The coupling to D1 breaks this perturbative shift symmetry down to integer shifts.⁵ This is evident from the tension formula (13.13) which is invariant under

⁵ This quantization of the shifts may be seen as a manifestation of Dirac quantization [13].

$$T: C^{(0)} \rightarrow C^{(0)} + 1, \quad (p, q) \rightarrow (p + q, q). \quad (13.18)$$

The integer shift $T: \tau \rightarrow \tau + 1$ together with the weak-strong duality $S: \tau \rightarrow -1/\tau$ generate the full modular group $SL(2, \mathbb{Z})$ (Sect. 5.1). Since T is a symmetry already in perturbation theory, S -duality implies the full $SL(2, \mathbb{Z})$ duality.

The IIB NS5-Brane

We need to understand how the $SL(2, \mathbb{Z})$ dualities act on the extended objects of Type IIB. We know that the (p, q) -strings form a single orbit of $SL(2, \mathbb{Z})$.

The 4-form RR potential $C^{(4)}$ is invariant under $SL(2, \mathbb{Z})$, and so $SL(2, \mathbb{Z})$ should take the D3-brane into itself. In particular, the D3-brane is *self-dual* under the strong-weak S -duality. The $SL(2, \mathbb{Z})$ duality acts on the background fields e^Φ , $C^{(0)}$ which look as Lagrangian couplings from the viewpoint of the 4d QFT on the D3 world-volume. The fact that IIB has a $SL(2, \mathbb{Z})$ then yields fundamental non-perturbative results for 4d QFT that we shall discuss in Sect. 13.1.2 below.

The D5-brane is a magnetic source for the RR 2-form charge: the integral of $F^{(3)}$ over a 3-sphere surrounding it is non-zero. Weak-strong duality should transform the D5 into a half-BPS magnetic source for the NS-NS 2-form gauge field $B_{\mu\nu}$. An object with precisely these properties was found in Sect. 8.8 as a classical half-BPS solution to the IIB field equation, i.e. as a (everywhere regular) soliton which we call the *NS 5-brane* or NS5 for short. The extremal solution is (see Eqs. (8.135), (8.137))

$$G_{mn} = e^{2\Phi} \delta_{mn}, \quad G_{\mu\nu} = \eta_{\mu\nu} \quad (13.19)$$

$$H_{mnp} = -\epsilon_{mnp}{}^q \partial_q \Phi, \quad e^{2\Phi} = e^{2\Phi(\infty)} + \frac{Q}{2\pi^2 r^2}. \quad (13.20)$$

Here the y^m ($m = 1, \dots, 4$) are transverse to the 5-brane, the x^μ are tangent to it, and $r^2 = y^m y^m$. This soliton is the magnetically charged object required by string duality. Its existence is an automatic consequence of the $SL(2, \mathbb{Z})$ symmetry of the SUGRA e.o.m. Consistency of the duality requires the product

$$\tau_{D1} \tau_{D5} = \pi/\kappa^2 \quad (13.21)$$

to be equal to

$$\tau_{F1} \tau_{NS5} \quad (13.22)$$

by the Dirac quantization (which determines the product of the charges) combined with the BPS condition (which relates the charges to the tensions). This yields

$$\tau_{NS5} = \frac{2\pi^2 \alpha'}{\kappa^2} = \frac{1}{(2\pi)^5 g^2 \alpha'^3}. \quad (13.23)$$

Note that the NS5 is *much heavier* than the D5 at weak coupling $g \ll 1$. In a sense, the NS5, which has a tension of order g^{-2} , is a “more non-perturbative” object than the D-branes which have tensions $O(g^{-1})$.

More generally, the $SL(2, \mathbb{Z})$ symmetry requires the presence of bound states of p NS5-branes with q D5-branes for all coprime integers (p, q) . The corresponding extremal black brane solitons are obtained by acting with the $SL(2, \mathbb{Z})$ symmetry of the IIB SUGRA e.o.m. to the solution (13.19)–(13.20). We conclude that for all coprime pair (p, q) we have precisely one ultrashort supermultiplet (2^8 states) of (p, q) 5-branes with the charges of p NS5's and q D5's.

Degrees of Freedom Living on the IIB NS5 Branes We wish to understand the effective 6d QFT which lives on the world-volume of the solitonic brane NS5. One can study this problem from the standpoint of IIB SUGRA: one looks for normalizable massless fluctuations around the extremal solution (13.19)–(13.20).

Exercise 13.1 Show that the bosonic normalizable massless modes are the normal fluctuations corresponding to translations in the orthogonal directions (the Goldstone of the broken translations) together with modes forming a 6d massless vector field.

In addition we have the fermionic partners of these bosons. All together the massless states form one vector supermultiplet of 6d (1, 1) SUSY. This result can be obtained without computations. By the $SL(2, \mathbb{Z})$ invariance of the IIB e.o.m., the normalizable massless fluctuations are the same for all (p, q) 5-branes, and hence equal to the light d.o.f. propagating on a single D5 which is an Abelian 6d (1, 1) vector supermultiplet.

Parallel NS5's As discussed in Sect. 8.8, the solution for any number of parallel NS5-branes is simply given by substituting

$$e^{2\Phi(y)} = e^{2\phi(\infty)} + \frac{1}{2\pi^2} \sum_{i=1}^N \frac{Q_i}{|y - y_i|^2}, \quad y \in \mathbb{R}^4 \quad (13.24)$$

into our earlier solution (13.19)–(13.20). This configuration preserves the same 16 supercharges as a single NS5, and hence there is no net force between parallel NS5 branes: this is required by S -duality since there is no force between parallel D5's. The low-energy field theory living on a stack of N IIB NS5's is 6d (1, 1) SYM with gauge group $U(N)$ just as for a stack of D5's.

Branes Ending on NS5 A fundamental string can end on a D5-brane. It follows by S -duality that a D-string should be able to end on a NS5-brane. Similarly, a D-string can end on a D3-brane. As in the case of F1 ending on a Dp -brane, there is a nontrivial constraint for one BPS object A to possibly end on a second object B : since A carries a conserved charge, the coupling between spacetime forms and world-brane fields living on B must allow B to carry the charge of A . Compare with our discussion of F1-D1 bound states in Sect. 12.7.1.

A D-string can stretch from one NS5 to another. The ground states of this D1 form a quarter-BPS supermultiplet. Indeed, such a state is related by S -duality to a ground state of the F-string extended between two D5s, which is related by T -duality to a massless open string in Type I theory. The mass of the D1 is given by the classical

D-string action in the above background and agrees with the prediction of S -duality. In particular, it vanishes as the NS5-branes become coincident, so like the D-branes they gain a non-Abelian symmetry in the limit. As already stated, they form the $(1, 1)$ SYM theory with $G = U(N)$, where N is the number of NS5's on top of each other.

Toward F -Theory

The $SL(2, \mathbb{Z})$ duality of Type IIB is the starting point for F -theory. F -theory is a deep quantum system (proposed by Cumrun Vafa [14]) which may be regarded as a far-reaching *non-perturbative completion* of Type IIB string theory. F -theory is a subject which deserves a full book on its own right; in this textbook we limit to present the general ideas beyond it as an invitation to the more advanced aspects of string theory (See Sect. 14.3).

13.1.2 D3-Branes and Montonen–Olive Duality

Consider a system of N parallel D3-branes at zero separation. The dynamics on their world-volume is 4d $\mathcal{N} = 4$ $U(N)$ SYM theory with gauge coupling (12.57) equal to

$$g_{\text{D3}}^2 = 4\pi g. \quad (13.25)$$

The YM coupling is dimensionless as it should be in 4d. At energy far below the Planck scale, $E \ll l_{\text{Pl}}^{-1}$, the coupling of the closed strings to the D3-brane modes becomes weak, and the D3-brane gauge theory decouples from the bulk string dynamics. The $SL(2, \mathbb{Z})$ duality of the IIB string takes this system to itself at a different coupling. In particular, at $\theta = 0$ the S -duality duality $g \rightarrow 1/g$ takes

$$g_{\text{D3}}^2 \rightarrow \frac{16\pi^2}{g_{\text{D3}}^2}. \quad (13.26)$$

This is a weak-strong duality transformation within the gauge theory itself. Thus, the self-duality of the IIB string implies a similar S -duality within 4d $\mathcal{N} = 4$ gauge theory. Such a duality was conjectured by Montonen and Olive in 1979 [15]. The evidence for the field-theoretic S -duality is of the same type as for the stringy S -duality: the BPS masses and degeneracies transform in the expected way as does the low energy effective action. People was skeptical about this conjecture, until string theory led to a derivation for it. In fact, as we shall see below, the string perspective gives us a purely geometric proof of the duality.

The field theoretic S -duality of 4d $\mathcal{N} = 4$ SYM extends to a full $SL(2, \mathbb{Z})$ duality as its stringy counterpart. The YM d.o.f. have a coupling (12.50) to the RR scalar

$$\frac{1}{4\pi} \int C^{(0)} \text{tr}(F^{(2)} \wedge F^{(2)}). \quad (13.27)$$

This is the Pontrjagin class ($\equiv$ instanton winding number) term of the YM action with $C^{(0)} = \theta/2\pi$. The full gauge theory action in a constant $C^{(0)}$ background is⁶

$$-\frac{1}{2g_{D3}^2} \int d^4x \operatorname{tr} |F^{(2)}|^2 + \frac{\theta}{8\pi^2} \int \operatorname{tr} (F^{(2)} \wedge F^{(2)}). \quad (13.28)$$

The T -duality $C^{(0)} \rightarrow C^{(0)} + 1$ is then the invariance under the shift

$$\theta \rightarrow \theta + 2\pi, \quad (13.29)$$

which arises from the quantization of the instanton charge. This shift invariance and the weak-strong duality (13.26) generate the full $SL(2, \mathbb{Z})$. It is more convenient to consider the complex YM coupling defined as

$$\tau \stackrel{\text{def}}{=} \frac{\theta}{2\pi} + \frac{4\pi i}{g_{D3}^2} = C^{(0)} + ie^\Phi \equiv \tau. \quad (13.30)$$

The $SL(2, \mathbb{Z})$ duality acts on the complex YM coupling τ by modular transformations. The instanton of Pontrjagin number k can be seen as a bound state of the stack of D3's with k D(-1)'s as discussed in Sect. 12.8.

Higgs Phase Let the D3-branes be parallel but slightly separated, corresponding to spontaneous breaking of $U(n)$ to $U(1)^n$. The ground state of an F-string stretched between D3-branes is BPS, and corresponds to a vector multiplet that has gotten mass from spontaneous breaking. The weak-strong dual is a D1-string stretched between D3-branes. To be precise, this is what it looks like when the separation of the D3-branes is large compared to the string scale. When the separation is small there is an alternative picture of this state as an 't Hooft–Polyakov magnetic monopole in the gauge theory [16–18]. The size of the monopole varies inversely with the energy scale of gauge symmetry breaking and so inversely with the separation. This is similar to the story of the instanton in previous chapter which has a D-brane description when it is small and a gauge description when it is large.

This is a first example of the interplay between spacetime dynamics of various branes and the non-perturbative dynamics of field theories which live on them.

13.2 *U*-Duality

The effect of toroidal compactification on the group of dualities is remarkable and instructive [6]. We consider the 10d Type IIB superstring compactified down to $d = 10 - k$ on a k -torus T^k . We have seen in Sect. 12.1 that when $k \geq 1$ this theory is equivalent to Type IIA compactified on a T -dual k -torus.

⁶ We write only the leading terms at low energy.

The low-energy theory is a 2-derivative, 32-supercharge, SUGRA in d dimensions. Since 32 is the maximal number of supercharges in any dimension (Sect. 8.3), we shall refer to such an effective theory as an *ungauged maximal supergravity in d dimensions*. The adjective *ungauged* means that in the IR description there is *no* light degree of freedom which is charged (electrically or magnetically) under a gauge form-field (including zero-form potentials as the RR Peccei–Quinn axion $C^{(0)}$). That the low-energy theory is *ungauged* follows from the fact that the states in a non-zero charge sector have masses bounded away from zero by virtue of the BPS bound.⁷

Facts about Maximal Supergravity

We already know that, for all d , the (ungauged) maximal SUGRA is *essentially uniquely* determined by supersymmetry alone (Chap. 8). Let us recall the fundamentals: there is a unique *covering* maximal SUGRA whose scalars' manifold $\widetilde{\mathcal{M}}$ is simply-connected.⁸ The manifold $\widetilde{\mathcal{M}}$, when equipped with the Riemannian metric $G(\phi)_{ij}$ which enters in the scalars' kinetic terms of the *Einstein frame* Lagrangian

$$L_E = -\frac{1}{2\kappa^2} \sqrt{-g} R + \cdots + \frac{1}{2} \sqrt{-g} G(\phi)_{ij} \partial^\mu \phi^i \partial_\mu \phi^j + \cdots, \quad (13.31)$$

is *globally* isometric to a rank $k + 1 \equiv 11 - d$ Riemannian symmetric space of non-compact type

$$\widetilde{\mathcal{M}}_d = G_d(\mathbb{R})/K_d \quad (13.32)$$

where $G_d(\mathbb{R})$ is a *connected, non-compact, real* Lie group and $K_d \subset G_d(\mathbb{R})$ is a maximal compact subgroup [19, 20] (unique modulo conjugacy in $G_d(\mathbb{R})$).

The *covering* d -dimensional maximal SUGRA admits $G_d(\mathbb{R}) \equiv \text{Iso}(\widetilde{\mathcal{M}}_d)$ as a group of *formal* symmetries acting on the scalars by isometries of the manifold $\widetilde{\mathcal{M}}_d$, and linearly on the gauge-invariant field strengths $\tilde{F}^{(s)}$ of the gauge forms. Here the adjective *formal* is used in the following sense: the SUGRA e.o.m. (but not necessarily the Lagrangian L_E [13]) admit $G_d(\mathbb{R})$ as a group of symmetries *provided* we restrict to topologically trivial field configurations. This restriction does not yield the actual effective theory: as discussed in Sect. 8.1 the physical effective theory contains sectors with non-trivial electro-magnetic fluxes which are integrally quantized *à la* Dirac whose allowed values form a certain self-dual⁹ lattice Λ . Since $G_d(\mathbb{R})$ acts on the field strengths, it acts on their fluxes, hence on $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$. Only the subgroup $\text{Aut}(\Lambda) \subset G_d(\mathbb{R})$ which preserves the lattice Λ is a bona fide group of symmetries of the low-energy theory. By definition, the subgroup of a semi-simple

⁷ Away from boundary points in moduli space where a *finite* or an *infinite* number of BPS states get massless and the low-energy description breaks down [1].

⁸ More precisely $\widetilde{\mathcal{M}}$ is the *covering moduli space*. We use the rough term “scalars' manifold” to refer loosely to the moduli space, i.e. the space parametrizing the physically inequivalent solutions to the equations of motion which are Poincaré invariant in d dimensions (a.k.a. “flat vacua”).

⁹ The term “self-dual” is here used in a slightly more general sense than elsewhere in this book. Λ may be either equipped with a symmetric or a skew-symmetric quadratic form. In both cases the quadratic form identifies $\Lambda \simeq \Lambda^\vee$.

Table 13.1 Group disintegration for maximal SUGRA in d dimensions

d	$G_d(\mathbb{R})$	K_d
10A	$SO(1, 1, \mathbb{R})$	1
10B	$SL(2, \mathbb{R})$	$SO(2, \mathbb{R})$
9	$SL(2, \mathbb{R}) \times SO(1, 1, \mathbb{R})$	$SO(2, \mathbb{R})$
8	$SL(2, \mathbb{R}) \times SL(3, \mathbb{R})$	$SO(2, \mathbb{R}) \times SO(3, \mathbb{R})$
7	$SL(5, \mathbb{R})$	$SO(5, \mathbb{R})$
6	$SO(5, 5, \mathbb{R})$	$SO(5, \mathbb{R}) \times SO(5, \mathbb{R})$
5	$E_{6(6)}(\mathbb{R})$	$Sp(4, \mathbb{R})$
4	$E_{7(7)}(\mathbb{R})$	$SU(8, \mathbb{R})$
3	$E_{8(8)}(\mathbb{R})$	$SO(16, \mathbb{R})$

real linear group¹⁰ which preserves a lattice Λ is an *arithmetic (sub)group* contained in (and commensurable to) a maximal arithmetic subgroup of $G_d(\mathbb{R})$ [21–23].

All other d -dimensional (ungauged) maximal supergravities (with the same Dirac lattice Λ) are obtained from the covering one by gauging a subgroup $\Gamma \subseteq \text{Aut}(\Lambda)$. Thus the (ungauged) maximal SUGRAs in d dimensions are classified by the pairs (Λ, Γ) where Λ is a charge lattice (up to equivalence) and $\Gamma \subseteq \text{Aut}(\Lambda)$ is a subgroup (modulo conjugacy). An effective maximal SUGRA has a consistent non-perturbative completion for a unique choice of (Λ, Γ) : the standard swampland conjectures [1] require that all fluxes allowed by Dirac quantization are realized (i.e. Λ is maximal) while $\Gamma \equiv \text{Aut}(\Lambda) \equiv G_d(\mathbb{Z})$, where $G_d(\mathbb{Z})$ is the subgroup of $G_d(\mathbb{R})$ of integral-valued elements. Pragmatically: $G_d(\mathbb{R})$ acts on $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$ by matrices; $G_d(\mathbb{Z})$ is the subgroup of elements represented by matrices with integral entries (in a basis of $\mathbb{Z}$ -generators of Λ).

The formal symmetry group $G_d(\mathbb{R})$ of the low energy maximal supergravity theory grows with the number k of compactified dimension as in Table 13.1; for proofs see [13]. In ref.[13], this list of groups is called *group disintegration* following [24].

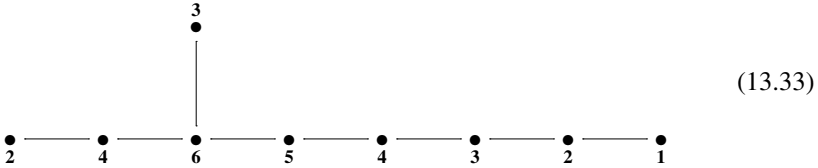
Remark 13.1 The real Lie groups $G_d(\mathbb{R})$ are all $\mathbb{R}$ -split real forms of the corresponding complex Lie groups.¹¹ When semi-simple, $G_d(\mathbb{R})$ is the real Lie group underlying the real locus of the Chevalley group-scheme [22, 25–29] of type G_d associated with the smallest super-lattice of the root lattice of G_d which contains the weights of the representation of $G_d(\mathbb{R})$ on $\Lambda \otimes_{\mathbb{Z}} \mathbb{R}$: this is the universal Chevalley group-scheme of the given type except for $d = 6$. $G_d(\mathbb{R})$ has a “God-given” arithmetic subgroup $G_d(\mathbb{Z})$, i.e. the “group of integral points” of the group-scheme which is the “canonical” arithmetic group required by quantum consistency (see, e.g. ref.[22] where the $d = 4$ is worked out in detail for Type II compactifications on T^6). These facts imply that $G_d(\mathbb{Z})$ enjoys a *maximality property* in the set of discrete subgroups of $G_d(\mathbb{R})$, that is, a discrete group Γ such that

¹⁰ Seen as an algebraic group defined over $\mathbb{R}$.

¹¹ This is already obvious from the fact that the rank of the symmetric space should be $k + 1$ by elementary KK considerations, while the rank of a non-compact symmetric space is the $\mathbb{R}$ -rank of its isometry group [20] which for the split form is equal to the (absolute) rank of the Lie group $G_d(\mathbb{R})$. The last equality is the defining property of the split form.

$G_d(\mathbb{Z}) \subseteq \Gamma \subset G_d(\mathbb{R})$ must be $\Gamma \equiv G_d(\mathbb{Z})$ [30]. The maximality property of Γ should be regarded as a *stronger version* of the spectral completeness conjecture of the swampland program [1, 31].

Let us describe the “group disintegration” rule which determines the groups $G_d(\mathbb{R})$ for the various d [13, 24]. The rule is best stated in terms of the Dynkin graphs¹² of $G_d(\mathbb{R})$. We start from the affine Dynkin graph of $\tilde{E}_8$ (cf. Fig. 2.3c)



which corresponds to the $d = 2$ “maximal SUGRA”,¹³ that is, to Type II with 8 compactified dimensions. Next we decompactify the dimensions one by one getting maximal SUGRA in the sequence of dimensions $d = 3, 4, \dots, 10$. Each time we decompactify a dimension we delete one node from the graph, starting from the extension node (the unique node with Dynkin label **1**, i.e. the rightmost one in (13.33)), getting the sequence of Dynkin graphs in table (13.34) which leads to the real Lie groups in Table 13.1

$d = 3$	$d = 4$	$d = 5$
$d = 6$	$d = 7$	$d = 8$

(13.34)

where, for the moment, we ignore the color of the nodes (white vs. black). The rule has a little ambiguity: it does not distinguish between $SL(2, \mathbb{R})$ and $SO(1, 1)$ —cf. the two inequivalent maximal SUGRAs in $d = 10$ in Table 13.1.

¹² The Dynkin graph of a simple real Lie group is the Dynkin graph of its *complexified* Lie algebra. Since we know that $G_d(\mathbb{R})$ is the split real form, the Dynkin graph of $G_d(\mathbb{R})$ determines uniquely the symmetric space $G_d(\mathbb{R})/K_d$.

¹³ 32-SUSYs $d = 2$ SUGRA is rather tricky, and we omitted this special case in our tables. However *graphically* $d = 2$ is a convenient point to start our “group disintegration”.

Table 13.2 The R-symmetry group K_d for maximal supergravity in d dimension may be read by the list of irreducible real Clifford modules M_d in view of the fact $DU(M_d^{\oplus 32/N(d)}) \subseteq K_d \subseteq U(M_d^{\oplus 32/N(d)})$.

d	3	4	5	6	7	8	9
$M_d^{\oplus 32/N(d)}$	$(\mathbb{R}^2)^{\oplus 16}$	$(\mathbb{C}^2)^{\oplus 8}$	$(\mathbb{H}^2)^{\oplus 4}$	$((\mathbb{H}^2)^{\oplus 2})^{\oplus 2}$	$(\mathbb{H}^4)^{\oplus 2}$	$(\mathbb{C}^8)^{\oplus 2}$	$(\mathbb{R}^{16})^{\otimes 2}$
$U(M_d^{\oplus 32/N(d)})$	$SO(16)$	$U(8)$	$Sp(4)$	$Sp(2) \times Sp(2)$	$Sp(2)$	$U(2)$	$SO(2)$

Note: $Sp(2) \sim SO(5)$ and $SU(2) \sim SO(3)$ (up to isogeny)

As for all SUGRAs with more than 16 SUSYs (cf. Note 8.2), the real form of $G_d(\mathbb{R})$ is uniquely determined by requiring that its maximal compact subgroup K_d is the R-symmetry group and hence satisfies¹⁴

$$DK_d = DU(M_d^{\oplus 32/N(d)}), \quad K_d \subseteq U(M_d^{\oplus 32/N(d)}), \quad (13.35)$$

where M_d stands for the irreducible real module of the *even* Clifford $\mathbb{R}$ -algebra $\text{Cl}(d-1, 1)^0$ (that is, the minimal $\mathbb{R}$ -spinor in Minkowski d -space) and $N(d) \equiv \dim_{\mathbb{R}} M_d$; cf. Eq. (8.23). This rule leads to Table 13.2 which agrees with Table 13.1 from group disintegration. In particular, we see that for maximal SUGRA $G_d(\mathbb{R})$ is the $\mathbb{R}$ -split real form of the group for all d .

Remark 13.2 The Satake diagram [32] of the $\mathbb{R}$ -split real form is just the Dynkin graph of the corresponding complex Lie algebra without automorphisms and with all nodes colored white.

Two Subgroups of $G_d(\mathbb{Z})$ with a Simple Physical Interpretation

We are familiar with two subgroups of the formal symmetry groups $G_d(\mathbb{R})$. The first is the $SL(2, \mathbb{R})$ formal symmetry of the uncompactified IIB theory. The second subgroup $O(k, k, \mathbb{R})$ is the formal symmetry of string compactified on T^k , the group which we encountered in Chap. 6 in the context of Narain compactifications. The embedding of these two subgroups in $G_d(\mathbb{R})$ is manifest from the decomposition of the Dynkin graphs (13.34) into the Dynkin subgraphs over the black (resp. white) nodes. From those graphs, it is clear that the $O(k, k, \mathbb{R})$ Dynkin graph is obtained from the $G_d(\mathbb{R})$ one by deleting the leftmost black node which is the Dynkin graph of the 10d subgroup $SL(2, \mathbb{R})$.

We know that, for both subgroups, the *actual* symmetry of the full string theory is the standard *arithmetic* subgroup: the $O(k, k, \mathbb{Z}) \subset O(k, k, \mathbb{R})$ T -duality group and, respectively, the $SL(2, \mathbb{Z}) \subset SL(2, \mathbb{R})$ S -duality group of 10d IIB. The continuous $O(k, k, \mathbb{R})$ is reduced to the discrete $O(k, k, \mathbb{Z})$ by the discrete spectrum of the (p_L, p_R) charges which take value in an even self-dual lattice of signature (k, k) , while the continuous $SL(2, \mathbb{R})$ is broken to $SL(2, \mathbb{Z})$ by the discrete spectrum of (p, q) -strings. These reductions of the symmetry are both manifestations of the general principle that the actual symmetry is the arithmetic subgroup of the formal

¹⁴ For any group G we denote by DG its *derived group*, i.e. the normal subgroup generated by its commutators. If G is a reductive Lie group (e.g. a compact Lie group) DG is its semi-simple part.

symmetry which preserves the Dirac lattice of charges/fluxes Λ . In the massless limit, the charged states (which are necessarily massive by the BPS bound) do not appear, and the symmetry appears to be continuous—but only as long as we limit ourselves to topologically trivial backgrounds: the flux sectors break the symmetry to the arithmetic group.

Note 13.2 It should be remarked that the two symmetry groups $G_d(\mathbb{R})$ and $G_d(\mathbb{Z})$ are *algebraically equivalent* in the sense that there does not exist an *algebraic* invariant which distinguishes them. In the physical language, there is no order parameter ϕ (e.g. the v.e.v. of a Higgs field), transforming in a finite-dimensional linear representation of $G_d(\mathbb{R})$, which distinguishes the two groups: all $G_d(\mathbb{Z})$ invariant tensors are automatically invariant for the full group $G_d(\mathbb{R})$.

$G_d(\mathbb{Z})$ as an Exact Gauge Symmetry

From the decomposition of the graphs (13.34) into the white versus black Dynkin subgraphs¹⁵ of $SO(k, k, \mathbb{R})$ and $SL(2, \mathbb{R})$, we see that the two non-commuting subgroups $SO(k, k, \mathbb{R})$ and $SL(2, \mathbb{R})$ generate the full group $G_d(\mathbb{R})$. Indeed, all the Chevalley–Serre generators [27, 33] of the Lie algebra $\mathfrak{g}_d(\mathbb{R})$ of $G_d(\mathbb{R})$ belong to the subspace

$$\mathfrak{so}(k, k, \mathbb{R}) + \mathfrak{sl}(2, \mathbb{R}) \subset \mathfrak{g}_d(\mathbb{R}), \quad (13.36)$$

so the connected components of the two groups $SO(k, k, \mathbb{R})$ and $SL(2, \mathbb{R})$ generate the full group $G_d(\mathbb{R})$ (since it is connected). Then

Lemma 13.1 *The duality subgroups $O(k, k; \mathbb{Z})$ and $SL(2, \mathbb{Z})$ generate the full arithmetic group $G_d(\mathbb{Z})$ (modulo finite groups¹⁶).*

We omit the proof: the idea is that the statement holds for the corresponding Chevalley groups. A weaker version of the **Lemma** is easy to get. Let $H \subset G_d(\mathbb{Z})$ be the subgroup generated by $SO(k, k; \mathbb{Z})$ and $SL(2, \mathbb{Z})$. The $\mathbb{R}$ -Zariski closure of $H^{\mathbb{R}} \subset G_d(\mathbb{R})$ is a Lie group which contains $SO(k, k, \mathbb{R})$ and $SL(2, \mathbb{R})$ hence the full $G_d(\mathbb{R})$. We conclude that the discrete group $H \subseteq G_d(\mathbb{Z})$ is *dense* in $G_d(\mathbb{R})$. By itself this does not imply that H is commensurable to $G_d(\mathbb{Z})$ (but it makes it to sound very plausible).

Since we already know that both T -duality and S -duality are *exact* symmetries of the full string theory, we conclude that, for all d , the maximal arithmetic subgroup $G_d(\mathbb{Z})$ of the low-energy formal symmetry $G_d(\mathbb{Z})$ is actually a symmetry of the full theory. In fact, since S - and T -dualities are *gauge* symmetries, so it must be $G_d(\mathbb{Z})$. The symmetry $G_d(\mathbb{Z})$ is called *U-duality* (the *universal* duality group [6]). The *U*-duality group $G_d(\mathbb{Z})$ is an exact gauge symmetry of Type II superstring compactified on T^{10-d} .

In perturbation theory we only see dualities which act linearly on the string coupling g , and so are symmetries of each term in the perturbative series—these are the

¹⁵ This decomposition holds even at the level of Satake diagrams [32].

¹⁶ We are cavalier about finite groups. However they are crucial for the existence of a consistent non-perturbative completion on the low-energy SUGRA.

T -dualities plus integral shifts of the RR fields. The other dualities take small g to large g and so require some understanding of the exact theory. The main tools we use are the **Tentative Physical Principle** and SUSY non-renormalization properties.

We can arrive to the same conclusion by a perhaps more physical argument. The (discrete) gauge group $\mathcal{G}_d$ must be a normal subgroup of the group of exact symmetries, $\mathcal{G}_d \triangleleft G_d(\mathbb{Z})$. $\mathcal{G}_d$ contains the T -duality group $SO(10-d, 10-d; \mathbb{Z})$ which was shown to be a gauged symmetry in Chap. 6. Hence $\mathcal{G}_d$ is an infinite normal subgroup of $G_d(\mathbb{Z})$. For $3 \leq d \leq 7$ we can use the deep theorem below, to conclude that $\mathcal{G}_d$ is of finite index in $G_d(\mathbb{Z})$, i.e. differs from $G_d(\mathbb{Z})$ by at most a finite group. The finite group $G_d(\mathbb{Z})/\mathcal{G}_d$, if non-trivial, is an unbroken global symmetry of the low-energy theory. On general grounds [1, 34] we expect that no such symmetry exists. Then $\mathcal{G}_d \equiv G_d(\mathbb{Z})$ (cf. **Tentative Physical Principle**).

Theorem 13.1 (Margulis normal subgroup theorem [23]) *Let $G(\mathbb{R})$ be a simple real Lie group with $\mathbb{R}$ -rank $G(\mathbb{R}) \geq 2$. Then a normal subgroup $N \triangleleft G(\mathbb{Z})$ is either finite or has finite index in $G(\mathbb{Z})$.*

Note 13.3 The volume of the moduli space of Type IIB compactified on T^{10-d}

$$\mathcal{M}_d = G_d(\mathbb{Z}) \backslash G_d(\mathbb{R}) / K_d \quad (13.37)$$

is *finite* see Box 6.10 for the precise value.

Example: Type II on T^5 Let us look at the example of the IIB string on T^5 which by T -duality is the same as the IIA string on T^5 . This example is chosen because it gives the simplest setting for Black Hole microstate counting (see Sect. 14.6). Let us count the gauge fields. From the NS-NS sector there are five KK gauge bosons $g_{\mu m}$ and five gauge bosons from the antisymmetric tensor $B_{\mu m}$. There are also 16 gauge bosons from the dimensional reduction of the various RR forms

$$5 \text{ from } C_{\mu n}, \quad 10 \text{ from } C_{\mu npq}, \quad 1 \text{ from } C_{\mu npqrs}. \quad (13.38)$$

The last vector $V^{(1)}$ is dual to the 2-form $C^{(2)}$, i.e. it is defined by the relation $dV^{(1)} = *_5 dC^{(2)}$. Finally $B_{\mu\nu}$ is dual to a vector in 5d. In total we have

$$\#_{\text{vectors}} = \overbrace{5+5+1}^{\text{NS-NS}} + \overbrace{5+10+1}^{\text{RR}} = 27, \quad (13.39)$$

gauge bosons. Recall that the fundamental representation of E_6 is the 27.

Let us see how T -duality group $SO(5, 5, \mathbb{Z})$ acts on the vectors. $SO(5, 5, \mathbb{Z})$ is generated by T -dualities along the various axes, linear rescalings of the axes, and discrete shifts of the antisymmetric tensor B_{mn} . This action rotates the first 10 NS-NS gauge fields (resp. the 16 RR gauge fields) between themselves and leaves the last NS-NS vector invariant. The finite-dimensional representations of $SO(5, 5, \mathbb{R})$ are just the analytic continuation of the finite-dimensional representations of $SO(10, \mathbb{R})$. Since $SO(5, 5, \mathbb{Z})$ is Zariski-dense in $SO(5, 5, \mathbb{R})$, the irreducible finite-dimensional

representations of the T -duality group are the restriction to the arithmetic subgroup of the analytic continuation of the irreducible finite-dimensional representations of the compact group $SO(10, \mathbb{R})$. The group $SO(10, \mathbb{R})$ has a vector representation **10**, two spinor representations **16_s** and **16_c**, and of course a singlet **1**. The gauge fields evidently transform in the corresponding representations of $SO(5, 5, \mathbb{Z})$. Which spinor representation occurs depends on whether we start with IIA or IIB which are flipped by an element of $O(5, 5, \mathbb{Z})$ of determinant -1 .

The low energy supergravity has a symmetry $E_{6(6)}$ which is a non-compact real form of E_6 . The arithmetic subgroup is denoted by $E_{6(6)}(\mathbb{Z})$. The group E_6 has a representation **27** and a subgroup $SO(10)$ under which

$$\mathbf{27} \rightarrow \mathbf{10} \oplus \mathbf{16} \oplus \mathbf{1}. \quad (13.40)$$

Evidently the gauge bosons transform in this **27**; this is, of course, already implied by 5d $\mathcal{N} = 4$ SUGRA. We identify the states carrying the various (gauge) charges (from the effective 5d perspective they look as ‘particles’ in $\mathbb{R}^{5,1}$). The charges transforming in the **10** are carried by the Kaluza–Klein and winding modes (i.e. F1’s wrapped on 1-cycles of T^5). Then U -duality also requires states carrying the charges in the **16**. These are just the various Dp-branes wrapped on cycles of T^5 . Finally the state carrying the **1** charge is a NS5-brane wrapped around the full¹⁷ T^5 .

U -Duality and BPS Bound States

Most of the bound states we found in Chap. 12 are actually required by U -duality, so their existence is a consequence of our **Tentative Physical Principle**.

By the arguments around Lemma 13.1 we may limit ourselves to U -dualities that can be written as a composition of T -dualities along a set $\{m, n, \dots, p\} \subset \{1, 2, \dots, 9\}$ of axes, which we shall denote as $T_{mn\dots p} \equiv T_m T_n \cdots T_p$, and the weak-to-strong S -duality S of Type IIB. $T_{mn\dots p}$ switches between Neumann and Dirichlet b.c. and between momentum and winding charge in the directions $\{m, n, \dots, p\}$. S interchanges the NS-NS and R-R 2-forms but leaves invariant the RR 4-form, and acts correspondingly on the objects carrying these charges. We write $D_{mn\dots p}$ for a D-brane extended in the directions $\{m, n, \dots, p\}$, and F_m a F-string extended along the m -axis (i.e. a unit winding charge in the m -th direction) and p_m a BPS state with a quantum of (compact) momentum in the m -th direction.

The first duality chain is¹⁸

$$(D_9, F_9) \xrightarrow{T_8} (D_{789}, F_9) \xrightarrow{S} (D_{789}, D_9) \xrightarrow{T_9} (D_{78}, D_\emptyset). \quad (13.41)$$

Thus the D1-F1 bound state is U -dual to the D0–D2 bound state. Both systems have precisely one ultrashort supermultiplet of BPS states, and the analysis of the two

¹⁷ More precisely: wrapped around the fundamental 5-cycle of T^5 .

¹⁸ Here and below the symbol (X, Y) stands (a bit abusively) for a bound state carrying the additive quantum numbers of the two BPS objects X and Y rather than the two objects themselves.

systems in the previous chapter looked quite similar. However to identify them one needs a chain of dualities which involves the non-perturbative step S .

The second chain is

$$\begin{aligned} (D_{6789}, D_6) &\xrightarrow{T_6} (D_{789}, D_6) \xrightarrow{S} (D_{789}, F_6) \xrightarrow{T_{6789}} \\ &\xrightarrow{T_{6789}} (D_6, p_6) \xrightarrow{S} (F_6, p_6). \end{aligned} \quad (13.42)$$

The bound states of n D0's and m D4's are thus U -dual to fundamental string states with momentum n and winding number m in one direction. Let us compare the degeneracies of BPS states in the two cases. For the winding string, the argument which led to (7.184) for the heterotic string shows that the BPS strings satisfy

$$(N, \tilde{N}) = \begin{cases} (mn, 0), & mn > 0, \\ (0, -mn), & mn < 0. \end{cases} \quad (13.43)$$

Here N and $\tilde{N}$ are the number of excitations above the massless ground state. We see that the BPS states have only left-moving or only right moving excitations. The generating function for the number of BPS states is the usual string partition function,

$$\text{Tr } q^N = 2^8 \prod_{k=1}^{\infty} \left(\frac{1+q^k}{1-q^k} \right)^8, \quad (13.44)$$

or the same with $N \leftrightarrow \tilde{N}$. The counting (13.44) is most easily done with the re-fermionized θ_α . In terms of the ψ^μ the GSO projection gives several terms, which simplify using the abstruse identity (or Riemann relations). The string degeneracy (13.44) precisely matches the degeneracies N_{mn} of the D0–D4 bound states found in the previous chapter using the cohomology of the Hilbert scheme (cf. Claim 12.5). Indeed, the formulae for the cohomology groups of the Hilbert scheme of points on a smooth projective surface were first conjectured by Vafa and Witten [35] on the basis of the string dualities and later confirmed by mathematicians as a theorem in Algebraic Geometry [36]. This is an example of the (several) “triumphs” of physical methods in pure mathematics.

13.3 ★ IIA on K3 is Dual to Heterotic on T^4

Another important application of the tentative principle is the duality between Type IIA compactified on a K3 surface and the heterotic string¹⁹ on T^4 [37–39]. In both cases, the low energy effective theory is a 6d (1, 1) SUGRA (16 supercharges). The covering SUGRA is then unique for a given YM gauge group G . To establish the equality of the low-energy theories, it suffices to show that

¹⁹ Which heterotic string, $SO(32)$ or $E_8 \times E_8$, is immaterial since the two are dual when compactified on T^4 ; cf. Sect. 7.7.1.

G , the flux lattice Λ , and the discrete gauge group Γ are the same ones. In both models we expect Λ to be the maximal lattice consistent with Dirac quantization, while $\Gamma = \text{Aut}(\Lambda)$, so the essential point is to check the equality of the gauge groups, an issue with some subtlety [37].

The moduli space $\mathcal{M}_4$ of heterotic compactified on T^4 was described in Sect. 7.7:

$$\mathcal{M}_4 = SO(4, 20; \mathbb{Z}) \backslash \widetilde{\mathcal{M}}_4 \quad \text{where} \quad \widetilde{\mathcal{M}}_4 = SO(4, 20; \mathbb{R}) / (SO(4) \times SO(20)). \quad (13.45)$$

The symmetric space

$$\mathbb{R} \times \widetilde{\mathcal{M}}_4 \quad (13.46)$$

is the *covering* scalars' space of 6d (1,1) SUGRA coupled to 20 vector-multiplets [13]. The factor $\mathbb{R}$ is parametrized by the *graviscalar* ϕ (the scalar in the SUSY multiplet of the graviton) which corresponds to the scalar already present in 10d (the dilaton). At a generic point in moduli space the gauge group is $U(1)^{20}$ times the $U(1)^4$ associated with the four graviphotons (cf. final remarks in Sect. 7.8). At special points (fixed by finite subgroups of $SO(4, 20; \mathbb{Z})$) the gauge group enhances to a non-Abelian Lie group of rank 20; the reader is referred to Sect. 7.7 for details.

Let us count light vectors for IIA on K3 at a *generic* point of its moduli space, i.e. for IIA in a background $\mathbb{R}^{5,1} \times K3$ with a generic hyperKähler metric and flat 2-form $B^{(2)}$ on the K3 factor. Since K3 is simply-connected, we have no KK vector from G_{MN} or B_{MN} : all vectors come from the RR sector. One RR vector A_μ is already present in 10d. The other vectors appear from the decomposition of the 10d 4-form field-strength in harmonic forms on K3

$$\widetilde{F}^{(4)} = *_6 F_0^{(2)} + \sum_{a=1}^{22} \omega^a \wedge F_a^{(2)}, \quad (13.47)$$

where $\{\omega_a\}$ is a basis of $H^2(K3, \mathbb{R}) \simeq \mathbb{R}^{22}$. In total we have 24 vectors: 4 graviphotons and 20 matter vectors. Thus at a *generic* point we have 6d (1, 1) SUGRA coupled to 20 vector multiplets, and the low-energy theory coincides with the one for the heterotic string on T^4 .

We can be more precise. Consider the lattice $\Lambda_{K3} \equiv H_*(K3, \mathbb{Z})$ endowed with the bilinear form given by the intersection pairing. We know from Sect. 11.1.3 that the lattice Λ_{K3} is even and self-dual of signature (4, 20), hence equivalent to $H^4 \oplus (-E_8) \oplus (-E_8)$ (cf. Sect. 7.5). We can identify Λ_{K3} with the Narain lattice Λ_{Nar} of heterotic on T^4 : this gives us a one-to-one identification between states on the two sides of the duality.

At a special point in moduli space, we expect that IIA on K3 has *the same* non-Abelian enhancement as the heterotic string at that point of $\mathcal{M}_4$. In the heterotic language, a massless vector associated with a root of the gauge group appears at a point in $\mathcal{M}_4$ where there is a solution to

$$p_R^2 = 0 \quad \text{and} \quad p_L^2 = 2 \quad \text{with} \quad (P_L, P_R) \in \Lambda_{\text{Nar}}. \quad (13.48)$$

We need to translate this condition in the language of IIA on K3 [37]. For simplicity we consider an easy case. Suppose that for some choice of its complex structure, our K3 surface contains a holomorphic curve C of genus 0. From the adjunction formula [40, 41], we have $[C] \cdot [C] = -2$ which yields the second condition in (13.48). The first condition reduces to $\int_C \omega = 0$, where ω is the complexified Kähler form in that complex structure. In other words, a point of enhanced gauge symmetry is a point in moduli space where the volume of an embedded S^2 which represents a non-trivial homology class goes to zero [37]. The vector field which gets massless at this special point has a simple interpretation: a D2 wrapped on a 2-cycle of K3 looks as a particle from the 6d perspective; its mass is proportional to the minimal volume of a surface representing the homology class; as this volume goes to zero, the vector gets massless, and the unbroken gauge symmetry enhances by reverse Higgs mechanism.

We shall recover and extend this result from a more intrinsic perspective, using the duality between strongly-coupled IIA and M-theory,²⁰ in the final remarks of Sect. 14.1.1. We refer the reader to that section for additional details.

Moduli of (4, 4) SCFT The result can be stated in a different way. From the viewpoint of 2d QFT, the moduli space $\mathcal{M}_4$ (with its canonical locally symmetric metric) should be the conformal manifold of an interacting 2d (4, 4) SCFT with Virasoro central charge $c = 6$ equipped with its Zamolodchikov metric. The conformal manifold can be determined directly, see [42] or [43] for a more general analysis valid for all c . One gets precisely the manifold $\mathcal{M}_4$ in Eq. (13.45).

Geometry The same result may be obtained directly from classical geometry see [43, 44]. The moduli space of *algebraic* K3's is

$$SO(2, 19; \mathbb{Z}) \backslash SO(2, 19; \mathbb{Z}) / (SO(2) \times SO(19)). \quad (13.49)$$

The space which parametrizes inequivalent hyperKähler structures on K3, i.e. the hyperKähler metrics modulo equivalence, is the bigger space [45, 46]

$$SO(1, 1) \times \left(SO(3, 19; \mathbb{Z}) \backslash SO(3, 19; \mathbb{Z}) / (SO(3) \times SO(19)) \right). \quad (13.50)$$

($SO(1, 1) \simeq \mathbb{R}$ is the overall scale of the metric). Adding the 22 $B^{(2)}$ deformations we get (13.45).

13.4 $SO(32)$ Type I-Heterotic Duality

Type I also has a half-BPS D1-brane which gets light in Planck units as $g \rightarrow \infty$, so the same strategy we used to understand the strong coupling limit of Type IIB works also for Type I: the D1 becomes a light string whose coupling gets weaker and weaker as $g \rightarrow \infty$, so its perturbative treatment becomes asymptotically exact. However, there are new aspects due to non-orientability. In Type I the only RR potentials surviving the Ω -projection are the 2-form, which couples electrically to the D1 and magnetically to the D5, and the non-dynamical 10-form which couples to the D9. This is consistent with the requirement for 16 unbroken supersymmetries in Type I—indeed the D1 and D5 both have $\#_{\text{ND}} = 4k$ relative to the spacetime filling D9's associated with the $SO(32)$ Chan–Paton labels and so, being invariant under 8 supersymmetries in presence of the D9's, are half-BPS in Type I $G = SO(32)$.

Consider again a long D1 wrapped in the 1-direction. The Type I D1 differs from the IIB one in two ways. First there is a projection onto Ω -even states. The $U(1)$ gauge field, with vertex operator $\partial_r X^\mu$ is projected out. The collective coordinates, with vertex operator $\partial_n X^\mu$, are kept since the normal derivative is even under a flip of the boundary orientation. Alternatively: the generalized doubling trick in Table 12.1 implies that the relation between the left- and right-moving parts of a DD scalar field has an extra minus compared to the NN scalar, that is, the action of Ω on the transverse X^i oscillators ($i = 2, \dots, 9$) has an additional -1 compared to the action

²⁰ See Sect. 13.5.

on the usual 9–9 strings. By superconformal symmetry, this extends to the ψ^μ ; in particular Ω on the R ground states is no longer -1 but acts as²¹

$$-\beta = -\exp[\pi i(s_1 + s_2 + s_3 + s_4)], \quad (13.51)$$

which gives an additional rotation by π in the four planes transverse to the D1. This projection removes from the fermionic 1–1 string states of the IIB D1 the left-moving $\mathbf{8}_s$ and leaves the right-moving $\mathbf{8}_c$ (cf. Eq. (13.1)).

The second difference is the inclusion of the 1–9 strings with one end on the D1 and one on a D9. The end on the D9 carries the Type I Chan–Paton index, so the 1–9 string states are vectors of $SO(32)$. These strings have $\#_{\text{ND}} = 8$ so that the zero-point energy (Eq. (12.79))

$$\frac{\#_{\text{ND}}}{8} - \frac{1}{2} = \frac{1}{2}. \quad (13.52)$$

is positive, and there are no massless states in the NS sector. The R ground states are massless as always (by world-sheet SUSY). Only ψ^0 and ψ^1 are periodic in the R sector (cf. Table 12.1), so their zero modes generate two states

$$|s_0; i\rangle, \quad (13.53)$$

where $s_0 = \pm 1/2$ and i is the CP index for the D9 end. One of these two states is removed by the GSO projection; with our choice of phases for the chirality operator

$$\exp(\pi i F) = -i \exp[\pi i(s_0 + s_1 + s_2 + s_3 + s_4)], \quad (13.54)$$

only the state with $s_0 = +1/2$ survives. We now impose the BRST condition, which as usual—cf. Eq. (13.2)—reduces to a Dirac equation, and then to the condition $s_0 = +\frac{1}{2}$ for the left-movers and $s_0 = -\frac{1}{2}$ for the right-movers. The right-moving 1–9 strings are thus removed from the massless spectrum by the combination of the BRST condition with the GSO projection. Finally we must impose the Ω projection; this determines the 9–1 state in terms of the 1–9 state, but gives no new constraint on the 1–9 states.

To summarize: the massless bosonic excitations are the usual collective coordinates X^i which describe the embedding of the D1-string in 10d space-time. The massless fermionic excitations are right-movers in the $\mathbf{8}_c$ of the transverse $SO(8)$ and left-movers which are vectors of the gauge $SO(32)$ and invariant under the transverse $SO(8)$. These are the same d.o.f. which live on the world-sheet of the $SO(32)$ heterotic string (as described in the *unitary* light-cone gauge).

Note 13.4 In Type I the D1 is a half-BPS object, i.e. invariant under 8 SUSYs. The 1–9 strings have massless R states (fermions) but *no* massless NS states (bosons). One may wonder how this is consistent with unbroken SUSY. The point is that these

²¹ Here β is the product of $\Gamma^m \Gamma$ on the D directions. The overall sign arises from our choice of conventions.

d.o.f. propagates only on the two-dimensional world-sheet, where we have a $(0, 8)$ SUSY: as in the heterotic string SUSY acts only on the right-movers, and the left-movers have no SUSY partner. This yields a check on our computation: $(0, 8)$ SUSY requires the 1-9 fermions to move in the opposite direction to the $1 - 1$ fermions. This is also required by the spacetime SUSY Ward identities; cf. Sect. 7.8.

The dependence of the D1 tension from the coupling

$$\tau_{\text{D1}} = \frac{1}{2\pi\alpha'g}, \quad (13.55)$$

is again exact. At strong coupling $\sqrt{\tau_{\text{D1}}}$ is the lowest mass scale in the theory, below the Planck mass and the F1 tension. By the same argument as in the IIB case, the natural conclusion is that the strongly coupled type I Theory is actually a weakly coupled $SO(32)$ heterotic string theory.

We check that this statement is consistent with low energy SUGRA, so the Type I/heterotic $SO(32)$ strong-weak coupling duality follows from our **Tentative Physical Principle**. In fact we have already noticed in Sect. 8.7 that the two effective theories are identical up to field redefinition, because they have the SUSY algebra and gauge group, and these two data determine uniquely the effective SUGRA (up to global issues²²) for 16 supercharges. It is important that the field redefinition we found in that section

$$G_{I\mu\nu} = e^{-\Phi_h} G_{h\mu\nu}, \quad \Phi_I = -\Phi_h \quad (13.56)$$

$$\tilde{F}_I^{(3)} = H_h^{(3)}, \quad A_I^{(1)} = A_h^{(1)}, \quad (13.57)$$

inverts the sign of the dilaton, i.e. switches strong $\leftrightarrow$ weak coupling.

We conclude that there is a single theory which looks like weakly coupled Type I theory when $e^{\Phi_I} \ll 1$ and like weakly coupled $SO(32)$ heterotic theory when $e^{\Phi_I} \gg 1$. At low energy Type I SUGRA is a good description of the physics for all values of Φ_I by SUSY non-renormalization. Indeed, even when the string coupling g is of order 1, the couplings in the low-energy theory are all irrelevant in 10d, and remain irrelevant as long as there are at least *five* non-compact dimensions,²³ and so are weak at low energy.

Going in the other direction, the Type I/heterotic duality also gives the strong-coupling physics of Type I: it is the weakly coupled $SO(32)$ heterotic theory. Note that the strategy we used above cannot be applied to the Type I fundamental string since it is not a BPS object and has no simple non-renormalization property. Indeed the Type I F -string does not carry any conserved charge: the NS-NS 2-form whose charge is carried by the IIB fundamental string is eliminated by the Ω -projection. The

²² The global aspects also agree between the two models, as a consequence of two basic principles: *spectral completeness* and *no ungauged symmetry*.

²³ The effective theory contains YM interaction which in four non-compact dimensions are strongly-coupled in the IR.

RR 2-form remains, but its charge is carried by the Type I D-string, not the F-string. That Type I F1 is not a BPS state is evident from the fact that it can break open. As the Type I coupling gets large the decay becomes more rapid, and as $g \rightarrow \infty$ the F1 ceases to be a meaningful object in the Type I theory.

We remarked in Chap. 7 that Type I and heterotic $SO(32)$ are perturbatively distinct theories since Type I string states carry only symmetric and antisymmetric 2-index representations of the gauge group $SO(32)$, whereas the states of the heterotic string come in arbitrarily large $SO(32)$ representations. These states appear in the Type I as D1 states, and one gets big representations by attaching to the D1 several 1–9 strings. In particular Type I D1 states can carry a spinor representation of $SO(32)$ which is not contained in any tensor product of 2-index representations. Let us describe the mechanism that makes this possible: we consider a D1 wrapped around a circle of length $L \gg \sqrt{\alpha'}$. The massless 1–9 strings are associated with fermions fields Λ^A living on the D1, with A the $SO(32)$ vector index. The zero modes of these fields

$$\Lambda_0^A = L^{-1/2} \int_0^L dx^1 \Lambda(x^1)^A, \quad (13.58)$$

satisfy a Clifford algebra

$$\{\Lambda_0^A, \Lambda_0^B\} = \delta^{AB}. \quad (13.59)$$

The quantization now proceeds just as for the fundamental heterotic string, producing in the left-moving R sector states which are spinors in the $2^{15}_s \oplus 2^{15}_c$ of $SO(32)$. The Λ^a are fields that create light 1–9 strings ending on our D1, and they play exactly the same role as the left-moving fermions λ^A that create excited states on the heterotic string.

$SO(32)$ GSO-Projection The attentive reader may be puzzled: as $g \rightarrow \infty$ the world-sheet Type I D1 d.o.f. look identical to the ones on the world-sheet of the $SO(32)$ heterotic string in the fermionic formulation, Chap. 7, *except* that in the heterotic case the physical left-moving states are *not* fermions in the **32** of $SO(32)$ but rather the NS and R states associated with the GSO projection of the $SO(32)$ current algebra associated with the even self-dual lattice Γ_{16} ; cf. Eq. (7.119). Quantum consistency (e.g. maximal locality of the operator algebra) requires that the left-moving fermions on the D1 world-sheet to be GSO-projected: there are two inequivalent consistent GSO-projections corresponding to the even self-dual lattices Γ_{16} and $\Gamma_8 \oplus \Gamma_8$ (cf. Sect. 7.6). Hence the Fermi d.o.f. on the D1 *must be* GSO-projected.²⁴ The existence of the GSO projection is not an issue. The only issue is *Which physical mechanism enforces the GSO projection on the light 1–9 strings attached to a Type I D1?*

To perform the GSO projection is equivalent to gauging a $\mathbb{Z}_2$ symmetry which acts as -1 on each 1–9 string endpoint lying on the D1. In the NS sectors the $\mathbb{Z}_2$ charge counts the Fermi fields $\Lambda_A \bmod 2$, while in the R sector it acts as the chirality

²⁴ If, by absurd, the left-moving Fermi d.o.f. on the world-sheet of the Type I D1 were *not* GSO projected, its world-sheet theory would give a *third* inequivalent SUSY heterotic string moving in 10d, a possibility which was rigorously ruled out in Chap. 7 in several ways.

operator of $Spin(32)$. The $\mathbb{Z}_2$ Gauss law projects the Hilbert space into the $\mathbb{Z}_2$ -invariant subspace: in the NS sector this keeps only operators which contain an even number of Λ^A 's, while in the R sector it keeps only one of the two $Spin(32)$ chiralities. To make the story more transparent, we compare this discrete gauging of the Type I D1 world-sheet theory with the corresponding gauging for the unprojected Type IIB D1. In the second case the D1 world-sheet model has a continuous $U(1)$ gauge symmetry and the F1 endpoints laying on the D1 carry a charge ± 1 (depending on the in/out orientation) under this $U(1)$. The $U(1)$ charge counts the number of Λ^A . The elements of the $U(1)$ gauge group which commute with the Ω projection, and so remain as gauge symmetries of the Type I D1, form the discrete subgroup $\mathbb{Z}_2 \subset U(1)$: it counts world-sheet fermions mod 2 and then its gauging produces the GSO projection.²⁵

String Tensions The tension of the Type I D-string is

$$\tau_{D1} \equiv \frac{1}{2\pi\alpha' g} = \frac{g_{YM}^2}{2^{1/2} 8\pi\kappa^2}, \quad (13.60)$$

where we used $\kappa^2 = 2^6\pi^7 g^2 \alpha'^4$ and Exercise 12.2 $g_{YM}^2 = 2 g_{D9}^2 \equiv 2^8\pi^7 g \alpha'^3$ to rewrite the tension in the form of a relation between three physical *observables* τ_{D1} , g_{YM} and κ which can be experimentally measured. The relation (13.60) is exact for all g by BPS non-renormalization. Type I/heterotic duality then requires the same relation to be satisfied by the (perturbative) heterotic string tension. This is indeed the case, as the reader is invited to check.

Note 13.5 The agreement of the string tensions on the two sides of the duality is not logically independent of our previous checks. Indeed the relation between the BPS string tension, the gravitational coupling κ , and the YM coupling g_{YM} can be read from the extremal solution of the effective 10d (1, 0) SUGRA with $G = SO(32)$; cf. Sect. 8.8. Since the effective SUGRA is the same on the two sides of the duality (modulo a field redefinition which cannot affect the observables) all relations between physical quantities which can be learned from the effective SUGRA are automatically consistent with the duality (this is the **Tentative Physical Principle** at work). In the literature, one finds many other such “quantitative checks” of the duality which are also “not logically independent”.

13.4.1 The Type I D5-Brane Versus the Heterotic NS5

The Type I D5-brane (the magnetic source of the RR 2-form field $C^{(2)}$) has some peculiar features which we quickly review. The D5–D9 system is T -dual to the D0–D4 system. In Sect. 12.8, we saw that in the second case the D0 is the zero-size limit of an instanton-particle of YM theory living on the D4 world-volume. Likewise, the

²⁵ Alternatively: implement the $U(1)$ Gauss law *before* taking the Ω projection.

D5 is an instantonic 5-brane of the space-time²⁶ Yang–Mills theory, namely a gauge field solution, whose fields are independent of the coordinates parallel to the D5, which looks as a Yang–Mills instanton in the 4 space directions orthogonal to the D5. This solution has collective coordinates for its position, and also for the size and gauge orientation of the instanton. In the instanton zero-size limit, the D5-brane description gets accurate. As in the discussion of the D0–D4 bound states, going to the Higgs branch we make the instanton bigger, and eventually the description in terms of classical solution to the Yang–Mills equations becomes good.

The $SO(32)$ heterotic theory should contain a half-BPS 5-brane dual to the Type I D5. This heterotic BPS brane is a magnetic source for the NS 2-form gauge field $B^{(2)}$ of the heterotic theory, so is a NS5 brane. As the Type IIB NS5, the heterotic NS5 is identified with a smooth soliton of the low energy effective theory: since the effective SUGRA is the same one on the two sides of the duality, and the Type I D5 soliton exists (Sect. 8.8), this implies that the heterotic NS5 soliton also exists. However the heterotic NS5 is *subtler* than its Type IIB sibling: by the identification of Type I D5 with instantons, we see that the relevant solitonic solution of SUGRA must have 10d gauge fields A^I with non-trivial topology (the magnetic NS charge of the 5-branes coincides with the Pontrjagin class of the 10d YM fields; cf. Sect. 12.8).

The classical SUGRA solution which describes a Type I D5 is easy to write. We denote x^μ (resp. y^i) the parallel (resp. transverse) coordinates to the 5-brane, and use again the *Ansatz* we introduced in Sect. 8.8 for the IIB D5 soliton (written in the string frame) which is mapped by S -duality to the *Ansatz* for the IIB NS5 soliton

$$ds_I^2 = e^{\Phi_I(y)} \eta_{\mu\nu} dx^\mu dx^\nu + e^{-\Phi_I(y)} dy^i dy^i, \quad (13.61)$$

$$\tilde{H}_{mnp}^{(3)} = \epsilon_{mnpk} \partial_k e^{-2\Phi_I(y)}. \quad (13.62)$$

Now there is a crucial modification with respect to Sect. 8.8. The Bianchi identity for the gauge-invariant 3-form field-strength $\tilde{H}^{(3)}$ of the NS 2-form $B^{(2)}$ is modified by the gauge and gravity Chern–Simons terms required by the Green–Schwarz mechanism

$$d\tilde{H}^{(3)} = cR^2 - c'F^2, \quad (13.63)$$

while the Yang–Mills topological charge

$$\frac{1}{8\pi^2} \int \text{tr } F^2 \quad (13.64)$$

is non-zero in presence of a YM instanton. Comparing with Eq. (13.62), we see that now $e^{-2\Phi_I(y)}$ is no longer a harmonic function since $d *_4 de^{-2\Phi} = cR^2 - c'F^2$ rather than $= 0$. Using the 't Hooft *Ansatz* [47] for the fields A_m of an instanton of size ϱ placed at the origin, we find [48, 49] (see also [50])

²⁶ Since the D9's are space-time filling the YM on their world-volume is just the 10d YM interaction of Type I.

$$e^{-2\Phi_I} = A + B \frac{r^2 + 2\varrho^2}{(r^2 + \varrho^2)^2} \quad A, B \text{ suitable constants} \quad (13.65)$$

(here $r^2 = y^m y^m$). Inserting this equation in (13.61) we get

$$\begin{aligned} ds_h^2 &= e^{-\Phi_I} ds_I^2 = \\ &= \eta_{\mu\nu} dx^\mu dx^\nu + \left(A + B \frac{r^2 + 2\varrho^2}{(r^2 + \varrho^2)^2} \right) (dr^2 + r^2 d\Omega^2). \end{aligned} \quad (13.66)$$

Note that in the zero-size limit $\varrho \rightarrow 0$ the solution (13.65), (13.66) reduces pointwise to the NS5 soliton configuration. From Eq. (13.66), we see that what looks in Type I like a small instanton becomes in the heterotic theory an instanton at the end of a long but finite throat of length

$$L_{\text{throat}} = \int_0^{r_0} \sqrt{G_{rr}} dr \approx -\sqrt{B} \log \varrho \quad \text{as } \varrho \rightarrow 0. \quad (13.67)$$

In the zero-size limit $\varrho \rightarrow 0$, the string frame throat becomes infinite as in the NS5 soliton in Fig. 8.1.

Symplectic Structure of Type I D5-Branes The Type I D5-branes have a second crucial difference with respect to the Type IIB D5's:

Claim 13.1 *In Type I one has to impose a symplectic rather than an orthogonal projection on the Chan–Paton d.o.f. of a set of n parallel D5's.*

We first describe some consequences of the **Claim**, deferring its justification to the end of this section. The massless bosonic states of the strings suspended between a set of parallel D5's are

$$\psi_{-1/2}^\mu |0, k; ij\rangle \lambda_{ij}, \quad \psi_{-1/2}^m |0, k; ij\rangle \lambda'_{ij}, \quad i, j = 1, \dots, 2n \quad (13.68)$$

where the $2n \times 2n$ matrices λ_{ij} (resp. λ'_{ij}) are the CP wave-functions of the D5 gauge fields (resp. normal coordinates) which are the D5 gauge fields and collective coordinates respectively. The symplectic Ω projection gives

$$\omega \lambda \omega^{-1} = -\lambda^t, \quad \omega \lambda' \omega^{-1} = \lambda'^t, \quad (13.69)$$

where $\omega \equiv i\sigma_2 \otimes \mathbf{1}_n$ is the standard $2n \times 2n$ symplectic matrix.

For $n = 1$ the general solution of (13.69) is

$$\lambda = \sigma^a, \quad \lambda' = 1. \quad (13.70)$$

The Chan–Paton wave-function for the brane collective coordinates is the identity, so $X_{ij}^m = X^m \delta_{ij}$, and the two would-be D5's corresponding to the two values of the CP index i are constrained to have the same shape and move together, and hence

form a single dynamic object which we interpret as a *single* Type I D5 brane which must be equipped with a Chan–Paton index taking two values. More generally, for n Type I D5's the CP index takes $2n$ values. This is similar to the T -dual of the Type I string, where the Chan–Paton takes 32 values but there are only 16 D-branes, each D-brane index being doubled to account for its orientifold image. The world-brane vectors have Chan–Paton wave-functions σ_{ij}^a so the gauge group living on a *single* D5 is $Sp(1) \simeq SU(2)$, contrary to the IIB D5-brane whose gauge group is $U(1)$.²⁷ For n coincident D5-branes the gauge group is $Sp(n)$ (simply-connected, compact real form of the Lie group of type C_n).

There are many ways to understand the symplectic nature of the Chan–Paton index for Type I D5's. The simplest argument is to consider the spectrum of 5-9 strings. For each value of the CP index there are two bosonic states as in Eq. (12.83). The D5–D9 system has eight SUSYs, and these two bosons form *half of a hypermultiplet*. In an oriented theory the 9–5 strings yield the other half hypermultiplet, and altogether we have a *full* hypermultiplet, but in the unoriented theory these are not new independent states, and we remain with just one *half* hypermultiplet. By definition, a half-hypermultiplet is a half-BPS supermultiplet of 8 supercharge SUSY, whose bosonic fields are all scalars, transforming in a symplectic (a.k.a. quaternionic) irreducible representation²⁸ of the gauge group G . We conclude that the world-volume gauge group must have quaternionic representation given by the CP vector space. The simplest such group is $SU(2)$ with the quaternionic representation $\mathbb{C}^2 \simeq \mathbb{H}$ given by the natural action as left multiplication by unit quaternions.

13.5 Type IIA at Strong Coupling: M-Theory

We continue with our strategy: to understand the strong coupling limit of a string theory we focus on the lightest mass scale associated with its BPS objects. If this scale is parametrically smaller than the Planck mass as $g \rightarrow \infty$, the states of the BPS object get weakly coupled and we can understand their dynamics in the limit. IIA has no BPS string, and the relevant d.o.f. do not describe a weakly coupled string, as it was the case for Types IIB and I. Type IIA at strong coupling is not a string theory: it should be a brand *new* quantum system.

²⁷ The two groups $Sp(1)$ and $U(1)$ are the groups of norm one elements in the division algebras (skew fields) $\mathbb{C}$ and $\mathbb{H}$, respectively. By Schur lemma the endomorphism ring of an irreducible representation is a skew-field, and the above groups are the corresponding compact automorphism groups.

²⁸ Recall that an irreducible $\mathbb{C}$ -representation V of a Lie group G is *real* (resp. *quaternionic*) iff $V \otimes V$ (resp. $V \wedge V$) contains the trivial representation [51]. If $V \otimes V$ does not contain the trivial representation, V is said *complex*. Let $\chi(g)$ be the character of V . The *Frobenius–Schur indicator* of V is $i(V) \equiv \int_G dg \chi(g^2)$ with dg the Haar invariant measure [52] normalized to $\int_G dg = 1$. The representation V is complex, real, or quaternionic iff $i(V)$ is 0, +1, or −1, respectively [53].

The D p -brane tension, $\tau_p = O(g^{-1}\alpha'^{-(p+1)/2})$, defines a mass scale

$$\Lambda_p \equiv (\tau_p)^{1/(p+1)} \approx g^{-1/(p+1)} \alpha'^{-1/2} \approx g^{(p-3)/(4(p+1))} M_{\text{Planck}}. \quad (13.71)$$

At strong coupling $g \gg 1$ the brane of smallest dimensions yields the lowest mass scale, and we must focus on the d.o.f. living on this lightest BPS object. For IIB this was the D1, and we got a dual string description; for IIA we must instead focus on the BPS 0-branes. There is an infinite sequence of them, corresponding to the bound states of D0's. The D0 mass is

$$\tau_0 = \frac{1}{g\alpha'^{1/2}}. \quad (13.72)$$

In the perturbative regime $g \ll 1$ this particle is very heavy, but it becomes light for $g \gg 1$. We know from Sect. 12.7.2 that any number n of D0's form precisely one $\frac{1}{2}$ -BPS supermultiplet of bound states. Their mass is fixed by the BPS condition to

$$n\tau_0 = \frac{n}{g\alpha'^{1/2}}. \quad (13.73)$$

This expression is *exact*. In the limit $g \rightarrow \infty$, the mass spectrum becomes continuum.

In IIB at strong coupling the spectrum of light BPS states was easily identified with the BPS states of a perturbative string. What about for the spectrum of light BPS states (13.73) in the limit $g \rightarrow \infty$? A continuum spectrum of particle states is characteristic of a non-compact quantum system. The mass spectrum (13.73) equals the mass spectrum of Kaluza–Klein states of a massless particle on a circle of radius

$$R_{10} = g\alpha'^{1/2}, \quad (13.74)$$

which have compact momentum $p_{10} = n/R_{10}$ with $n \in \mathbb{Z}$. In the limit $g \rightarrow \infty$, $R_{10} \rightarrow \infty$ and the circle decompactifies into a line. We are forced to conclude that in IIA the strong coupling limit $g \rightarrow \infty$ is a *decompactification* limit where spacetime acquires a new *extra* dimension. We shall see momentarily that the resulting theory in eleven non-compact dimensions has 11d Poincaré (super)symmetry, so the new dimension is a genuine one, geometrically unified with the original 10 dimensions. The non-perturbative completion of the 10d IIA theory lives in a 11d spacetime! This is one of the most surprising phenomena in string theory. This is yet another hint that the geometry of spacetime is a subtle notion in string theory (as it should be expected for a quantum theory of gravity).

According to our **Tentative Physical Principle**, this property should be already visible from the low-energy effective theory. We saw in Sect. 8.3 that 11d SUGRA is the supersymmetric field theory with the largest possible Poincaré invariance. For $d > 11$ spinors have ≥ 64 components, and this would lead to massless fields with spins greater than 2 which do not have consistent interactions. In Chap. 8 we used dimensional reduction of 11d SUGRA as a mere formal trick to get 10d IIA SUGRA. Now we understand that it was more than a trick: dimensional reduction keeps only

the $p_{10} = 0$ states, but string theory has natural candidates for states with $p_{10} \neq 0$ namely the bound states of D0's. To fully justify the identification of the 11d KK modes with D0 bound states, we have to show that the RR charge carried by the D0's is the KK electric charge or, dually, that the RR gauge vector A_μ which couples to the D0 charge is the KK vector. This was already shown in Chap. 8: the 10d RR vector A_μ of IIA arises from the dimensional reduction of the 11d metric, $A_\mu \approx G_{\mu 10}$, so it couples to the charge p_{10} (see discussion in Sect. 6.1). The D0 charge and p_{10} momentum are one and the same quantity in the effective SUGRA, and the **Tentative Physical Principle** extends this identification to the full non-perturbative completion. The non-perturbative 11d space-time of IIA is a principal $U(1)$ -bundle over the 10d space-time visible in perturbation theory, see Sect. 6.1. We conclude:

The strong coupling limit of IIA is an 11d SUSY quantum system known as **M-theory**. Its low-energy effective theory is 11d SUGRA

M-Theory Parameters

The 11d gravitational coupling is given by dimensional reduction as

$$\kappa_{11}^2 = 2\pi R_{10} \kappa^2 = \frac{1}{2} (2\pi)^8 g^3 \alpha'^{9/2}. \quad (13.75)$$

It is convenient to define the 11-dimensional Planck mass M_{11} to be

$$M_{11} = g^{-1/3} / \alpha'^{1/2}, \quad 2\kappa_{11}^2 = (2\pi)^8 M_{11}^{-9}. \quad (13.76)$$

The two parameters of the IIA theory, g and α' , are related to the 11d Planck mass M_{11} and radius of compactification R_{10} by Eqs. (13.74), (13.76); inverting

$$g = (M_{11} R_{10})^{3/2}, \quad \alpha' = M_{11}^{-3} R_{10}^{-1}. \quad (13.77)$$

M -theory has no dimensionless parameter which we may take small to produce an asymptotic expansion. However at energies $\ll M_{11}$ 11d SUGRA yields a good approximation to M -theory. At higher energies supergravity ceases to be reliable. To cover this regime one needs a non-perturbative definition of the mysterious M -theory. We shall comment on some possible definitions in Chap. 14.

Type IIB S -Duality $\Rightarrow$ M-Theory

We know the strong coupling limit of IIB: by S -duality, it is again weakly coupled IIB in a different $SL(2, \mathbb{Z})$ frame. Type IIA is T -dual to Type IIB: we must be able to deduce the strong coupling limit of IIA directly from the known one for IIB.

We compactify Type IIB on S^1 by identifying periodically the 9-direction. The IIB weak-strong duality S interchanges a D1 wound in the 9-direction (carrying one unit of the RR charge r_9) with an F1 string in the same direction. Under T -duality,

the D1 becomes a D0 and the wound F1 becomes a F1 with one unit of p_9 . TST then takes the D0 charge to p_9 and vice versa:

$$D0 \xrightarrow{T_9} (D1, r_9) \xrightarrow{S} (F, w_9) \xrightarrow{T_9} (F, p_9) \quad (13.78)$$

$$(F, p_9) \xrightarrow{T_9} (F, w_9) \xrightarrow{S} (D1, r_9) \xrightarrow{T_9} D0. \quad (13.79)$$

The chain of dualities identifies the D0 charge with the p_9 momentum of the dual theory, in agreement with our previous findings. We stress that, while the IIA and IIB strings are quite similar in perturbation theory, their strongly coupled behaviors look quite different. At strong coupling IIB is dual to itself, while IIA gets an extra spacetime dimension. Nevertheless these radically different behaviors are consistent with the equivalence of IIA and IIB under T -duality.

U-Duality versus M-Theory

For Type II on a circle the U -duality group U is

$$d = 9: \quad U = SL(2, \mathbb{Z}). \quad (13.80)$$

Regarded as a compactification of IIB, U is just the $SL(2, \mathbb{Z})$ duality of the 10d theory. Seen as a compactification of the IIA on a circle, and therefore of M-theory on T^2 , $SL(2, \mathbb{Z})$ is a geometric symmetry namely the mapping class group of the space-time factor T^2 . This identification is obvious when we regard the effective $d = 9$ maximal supergravity as the KK compactification of 11d SUGRA on T^2 .

For Type II on T^2 , the U -duality group is

$$d = 8: \quad SL(3, \mathbb{Z}) \times SL(2, \mathbb{Z}). \quad (13.81)$$

The T -duality subgroup for strings on T^2 is $SL(2, \mathbb{Z}) \times SL(2, \mathbb{Z})$, see Sect. 6.5. The first factor is the modular group of T^2 . The second factor in the Narain T -duality group may be understood as follows: flat T^2 is a Calabi–Yau of complex dimension 1. The conformal manifold of the σ -model with target T^2 is then

$$\mathcal{M}(T^2) \times \mathcal{M}(\text{mirror } T^2) = SL(2, \mathbb{Z}) \backslash \mathcal{H} \times SL(2, \mathbb{Z}) \backslash \mathcal{H}. \quad (13.82)$$

In the full U -duality group the mapping class group of T^2 is enlarged to the mapping class group $SL(3, \mathbb{Z})$ of the M-theory internal space T^3 , as expected. The “mirror” $SL(2, \mathbb{Z})$ is not modified since switching T^2 with its mirror switches from the strongly coupled IIA (i.e. M-theory) to strongly coupled IIB which is still 10-dimensional.

More generally, for Type II on T^{10-d} we look at the Dynkin diagram of the real Lie group $G_d(\mathbb{R})$, see Table 13.1, and consider the following Dynkin graph decomposition into subgraphs over the star versus circle nodes, which is alternative to our previous black/white node decomposition in (13.34),

$d = 3$	$d = 4$	$d = 5$	(13.83)
$d = 6$	$d = 7$	$d = 8$	

The decomposition shows that the U -duality group $G_d(\mathbb{Z})$ contains the geometric mapping class group $SL(11-d, \mathbb{Z})$ of the M-theory internal space T^{11-d} whose Dynkin diagram is the graph over the circle nodes. The full U -duality group is generated by this geometric M-theory group and the (non-commuting for $d \leq 7$) stringy $SL(2, \mathbb{Z})$ duality.

13.6 M-Theory BPS Objects Versus IIA Branes

Type IIA has a rich family of extended BPS objects. It is interesting to see how each one of them originates from M-theory on a circle [54] .

M2 and M5 Branes

We consider first the extended BPS objects of the 11d M-theory. There is one light tensor gauge field: the 3-form $B_{\mu\nu\rho}$. The corresponding electrically charged object is a 2-brane which we denote as M2. In the literature 2-branes are simply called *membranes*. The magnetically charged dual object is a 5-brane which we call M5. Of course the notions of *electric* and *magnetic* interchanges if we use the dual 6-form potential. However in d=11 SUGRA the formulation with a 3-form potential should be preferred, since the Chern–Simons coupling in its Lagrangian (8.28) has no local expression in terms of the dual 6-form.

We saw in Sect. 8.8 that 11d SUGRA has supersymmetric solutions carrying the appropriate charges to be identified with the M2 and M5 branes. Both of them preserve 16 supersymmetries.

Now we consider each Type IIA brane in turn.

IIA 0-Branes

The D0 branes of IIA are the BPS states with non-zero p_{10} charge. In M-theory they are states of the massless graviton supermultiplet, an ultrashort multiplet with 2^8 states, with a momentum $p_{10} \equiv n/R_{10}$ along the M-theory circle.

IIA 1-Branes

The F1 of IIA originates as an M2 brane wrapped on the extra circle. Indeed the wrapped M2 is the electric source of $C_{\mu\nu 10}$ which reduces in 10d to the NS-NS field

$B_{\mu\nu}$ whose electric source is the IIA string. The classical world-volume action of a wrapped M2-brane reduces to that of the Type IIA string [55].

IIA 2-Branes

The obvious origin of the IIA D2 is as a transverse (rather than wrapped) M2. A D2 couples to the RR potential $C_{\mu\nu\rho}$ which is the reduction of the 11d 3-form to which the M2 couples. When written in terms of the M-theory parameters, the D2 tension

$$\tau_{D2} = \frac{1}{(2\pi)^2 g \alpha'^{3/2}} = \frac{M_{11}^3}{(2\pi)^2} \quad (13.84)$$

depends only on the 11d fundamental scale M_{11} , and not on R_{10} , as required for an object that exists in the 11d theory. On the other hand, the F1 tension

$$\tau_{F1} = \frac{1}{2\pi\alpha'} = (2\pi)(g\alpha'^{1/2})\tau_{D2} \equiv 2\pi R_{10} \tau_{D2} \quad (13.85)$$

is linear in R_{10} as it should be for an object wrapped on the M-theory circle.

Membrane Magics The D2-brane is perpendicular to the extra direction, and should be free to move and fluctuate in that new direction. That is, the 3d QFT living on the world-volume of D2 should contain, in addition to the usual normal scalars X^i ($i = 3, \dots, 9$) an extra scalar field X^{10} , to be identified with the additional normal coordinate, so that the 3d QFT describes the membrane moving in *eleven* dimension not just in ten. This may look puzzling, because D-branes have collective coordinates only for their motion in the 10d spacetime of perturbative string theory. Consistency of IIA/M-theory duality requires the D2 world-volume QFT to have two (apparently) conflicting interpretations: one describing the motions of the membrane in 10 dimensions, and one describing its motions in 11 dimensions. The X^i 's are Goldstone bosons of broken translational symmetries, and the number of such bosons should be *both* ten and eleven.

The requirement that a 3d QFT must have two different interpretations, with inequivalent spontaneously broken symmetries, looks quite unlikely at first, and only a major miracle may save the day. This is the case: in 3d a vector field describes the same (local) physics as a scalar by Poincaré duality, and the required miracle indeed happens. We work out the details. The bosonic action for a D2 in flat space-time is

$$- \tau_2 \int d^3x \left\{ \left[-\det(\eta_{\mu\nu} + \partial_\mu X^m \partial X^m_\nu + 2\pi\alpha' F_{\mu\nu}) \right]^{1/2} + \frac{1}{2} \epsilon^{\mu\nu\rho} \lambda \partial_\mu F_{\nu\rho} \right\}. \quad (13.86)$$

Here we treat $F_{\mu\nu}$ as an independent field, so we included a Lagrange multiplier λ to enforce the Bianchi identity. $F_{\mu\nu}$ is then an auxiliary field (its e.o.m. determines it as a local function of the other fields) and it can be eliminated with the result

$$- \tau_2 \int d^3x \left\{ -\det[\eta_{\mu\nu} + \partial_\mu X^m \partial X^m_\nu + (2\pi\alpha')^{-2} \partial_\mu \lambda \partial_\nu \lambda] \right\}^{1/2} \quad (13.87)$$

see Box 13.2. Now we define

$$X^{10} = \lambda / (2\pi\alpha') \quad (13.88)$$

to write the action in the symmetric form

$$S[X^M] = -\tau_2 \int d^3x \left\{ -\det[\eta_{\mu\nu} + \partial_\mu X^M \partial_\nu X^M] \right\}^{1/2} \quad (13.89)$$

which is exactly the Nambu–Goto action for a membrane moving in eleven flat dimensions. We stress that the full 11d Poincaré invariance is restored. This is the M2-brane theory in its full glory.

More Membrane Magics The 3d QFT yields a much more precise description of the physics. The Lagrange multiplier field enforcing the Bianchi identity, λ , takes values in $\mathbb{R}$ if and only if the dual (magnetic) gauge group G_{mag} associated with $F_{\mu\nu}$ is *non-compact*, i.e. iff $G_{\text{mag}} \simeq \mathbb{R}$.²⁹ When the gauge group G is *compact*, i.e. $G \simeq G_{\text{mag}} \equiv U(1) \simeq S^1$, λ takes value in the group S^1 , i.e. it is a *periodic* scalar field³⁰ as we shall show momentarily. On the D2-brane there are objects carrying electric charge ± 1 , namely the endpoints of the fundamental string, so $G_{\text{mag}} \not\simeq \mathbb{R}$ (cf. Footnote 29). Even more crucially: we know that a D2 with k units of magnetic flux is a state carrying k units of D0 charge (cf. Eq. (12.191)); hence the magnetic flux on the D2 is quantized. By the Dirac argument (Sects. 8.8 and 12.4) this implies that the 3d electric charge is also quantized in integral units, and the magnetic D2 gauge group G_{mag} is $U(1)$ not $\mathbb{R}$.

The difference between a non-compact and a compact magnetic gauge group is that in the first case the 2-form field strength $F^{(2)}$ is (off-shell) just a *closed* 2-form.³¹ In the compact case $F^{(2)}/2\pi$ is a closed 2-form whose cohomology class is integral, since it represents the first Chern class of the associated $U(1)$ bundle (Dirac quantization). Consider the Lagrange multiplier/Chern–Simons term in the action (13.86) which enforces these conditions for $F^{(2)}$

$$S = \dots + \int (dB^{(0)}) \wedge F^{(2)}. \quad (13.90)$$

By Hodge decomposition [40], to enforce the above conditions on $F^{(2)}$, in the non-compact case, $dB^{(0)}$ should be an *exact* 1-form, i.e. $B^{(0)}$ is a globally defined scalar

²⁹ By $G_{\text{mag}} \simeq \mathbb{R}$ we mean that the magnetic charge is non-trivial and non-quantized, which by Dirac duality means that the electric charge is identically zero. Dually, $G \simeq \mathbb{R}$ means that the electric charge is non-quantized and the magnetic one trivial. In the compact case $G \simeq U(1)$ both electric and magnetic charges are non-trivial and integrally quantized.

³⁰ This follows from the Poisson resummation of the topological sectors in the 3d path integral (analogous to the 2d path integral discussed in Sect. 6.2). For more details see Sect. 1.6.1 of [13].

³¹ Dually, for a non-compact *electric* gauge group $F^{(2)}$ is an *exact* 2-form. To implement this dual condition the Lagrange-multiplier 1-form $dB^{(0)}$ in Eq. (13.90) should be an *arbitrary* closed 1-form representing a general class in $H^1(V_{D2}, \mathbb{R})$.

BOX 13.2 Proof of Eq. (13.87)

We work in Euclidean signature. We set $(a, b = 1, 2, 3, m = 3, \dots, 9)$

$$g_{ab} = g_{ba} = \delta_{ab} + \partial_a X^m \partial_b X^m, \quad f_{ab} = -f_{ba} = 2\pi\alpha' F_{ab}, \quad t^a = \frac{1}{2}\epsilon^{abc} f_{bc}$$

$$|t|^2 = g_{ab} t^a t^b, \quad \xi_a = \partial_a \lambda / (2\pi\alpha'), \quad |\xi|^2 = g^{ab} \xi_a \xi_b,$$

where g^{ab} is the inverse to g_{ab} . The Lagrangian reads (up to an overall constant)

$$[\det(g + f)]^{1/2} - it^a \xi_a.$$

One has the identity $\det(g + f) = \det g + |t|^2$, so the Lagrangian may be written as

$$\mathcal{L}(t, \xi) = \sqrt{\det g + |t|^2} - it \cdot \xi, \quad |t|^2 \equiv g_{ab} t^a t^b$$

. The equation of motion for t^a is

$$0 = \frac{\partial \mathcal{L}}{\partial t^a} = \frac{g_{ab} t^b}{\sqrt{\det g + |t|^2}} - i \xi_a$$

from which we get

$$-|\xi|^2 = \frac{|t|^2}{\det g + |t|^2} \Rightarrow |t|^2 = -\det g \frac{|\xi|^2}{1 + |\xi|^2} \Rightarrow \det g + |t|^2 = \frac{\det g}{1 + |\xi|^2}$$

. On the other hand,

$$it \cdot \xi = \frac{|t|^2}{\sqrt{\det g + |t|^2}} = (\det g)^{-1/2} |t|^2 (1 + |\xi|^2)^{1/2} = -(\det g)^{1/2} \frac{|\xi|^2}{\sqrt{1 + |\xi|^2}},$$

and then

$$\mathcal{L} \Big|_{\partial \mathcal{L} / \partial t^a = 0} = \frac{\sqrt{\det g}}{\sqrt{1 + |\xi|^2}} (1 + |\xi|^2) \equiv \sqrt{\det g (1 + |\xi|^2)}.$$

Finally we have the identity $\det(g_{ab} + \xi_a \xi_b) = \det g (1 + |\xi|^2)$. Returning to the original notation, we get

$$\mathcal{L} \Big|_{\partial \mathcal{L} / \partial F_{ab} = 0} = \left[\det(\delta_{ab} + \partial_a X^m \partial_b X^m + (2\pi\alpha')^{-2} \partial_a \lambda \partial_b \lambda) \right]^{1/2}$$

(thus taking value in $\mathbb{R}$). In the *compact* case $dB^{(0)}$ is a closed 1-form whose cohomology class is integral, that is, for all *closed* curve γ in the brane world-volume

$$\int_{\gamma} dB^{(0)} = n \in \mathbb{Z}. \quad (13.91)$$

This means that the scalar $B^{(0)}$ is well defined only up to shifts by integers, i.e. $B^{(0)}$ is a *periodic coordinate* on a circle of unit length

$$B^{(0)} \sim B^{(0)} + 1. \quad (13.92)$$

Comparing Eqs. (13.86), (13.90) we have

$$B^{(0)} = \tau_2 \lambda = 2\pi \alpha' \tau_2 X^{10}, \quad (13.93)$$

since $\tau_2 = [g (2\pi)^2 \alpha'^{3/2}]^{-1}$ and $R_{10} = g \alpha'^{1/2}$, we get

$$B^{(0)} = \frac{X^{10}}{2\pi R_{10}}, \quad (13.94)$$

and the Dirac quantization on the D2-brane implies that the 11-th coordinate X^{10} (of the space in which the dual M2-brane moves) satisfies the periodicity condition

$$X^{10} \sim X^{10} + 2\pi R_{10} \quad (13.95)$$

i.e. it is compactified on a circle of radius precisely R_{10} . Thus the D2 world-volume QFT already contains the information that the actual spacetime has the form $\mathbb{R}^{9,1} \times S^1$ where the circle has length $2\pi g \alpha'^{1/2}$. The M2-brane may be seen as the fundamental object of M-theory, and the world-volume theory on the M2-brane plays a role (in a sense) akin to the world-sheet SCFT for string theory. Having recovered from the world-volume action of the D2-brane the dynamics of a supersymmetric brane moving in $\mathbb{R}^{9,1} \times S^1$ means that (in principle) we have recovered the full 11d M-theory from this 3d QFT. It is remarkable that the spacetime geometry follows directly from 3d Dirac quantization. This D2 “miracle” gives us considerable confidence in the validity of the IIA/M-theory duality. The analysis may be extended to the fermionic terms (determined by SUSY) as well as to membranes moving in background fields and stacks of n M2’s. We add a couple of clarifying remarks.

Note 13.6 The scalars living on the world-volume of the D2/M2 are Goldstone bosons of the broken continuous symmetries. As always, we can reconstruct the broken symmetry group $\mathcal{G}$ from the low-energy physics of the Goldstones. What the above analysis says is that the broken symmetry group is actually $\mathbb{R}^7 \times U(1)$ with one compact factor (translations along a circle).

Note 13.7 It is clear from Eq. (13.86) that the 1-form $d\lambda$ is (up to normalization) the *magnetic dual field strength* of the electric field strength $F^{(2)}$. This explains why we had to discuss the physics in terms of the dual magnetic gauge group instead of the ordinary one.

IIA 4-Branes

These are wrapped M5-branes. Let us check their tensions

$$\tau_4 \equiv \frac{1}{g (2\pi)^2 \alpha'^3} = (2\pi g \alpha'^{1/2}) \frac{1}{(2\pi)^5 g^2 \alpha'^3} = (2\pi R_{10}) \tau_{\text{NS5}}. \quad (13.96)$$

IIA 5-Branes

Type IIA like Type IIB has a 5-brane object carrying the magnetic NS-NS $B_{\mu\nu}$ charge. The classical soliton is the same one as in the IIB theory, because the action for the NS-NS fields is the same in the two Type II SUGRAs, while the RR background is trivial in this solution. However, there are crucial differences. A D1 can end on the IIB NS5-brane. Performing T -duality in a direction parallel to the 5-brane, we obtain a D2 which ends on a IIA NS5-brane. From the point of view of the $(5 + 1)$ -dimensional field theory on the 5-brane, the end of a IIB D1 looks as a point-particle which is an electric source for the $U(1)$ gauge field living on the 5-brane. This is necessary³² so that the 5-brane, through a Chern–Simons interaction, can carry the RR charge of the D1. Similarly, the end of the IIA D2 is a *string* in the 5-brane, and it should couple to a *2-form gauge field* living on the world-volume of the IIA NS5-brane.

We were not surprised to find a $U(1)$ gauge field living on the IIB NS5-brane because it is related by S -duality to the IIB D5-brane which has such a world-volume field. We cannot use this argument to guess the effective field theory on the IIA NS5-brane. However in both cases the light fields living on the world-volume can be determined directly by looking to small fluctuations around the soliton classical solution. We summarize the results of the analysis; later we shall show how they can be predicted using (super)symmetry without any detailed computation.

Modes that are normalizable in the directions transverse to the 5-brane correspond to d.o.f. which live on the 5-brane. The bosonic d.o.f. include the collective coordinates for its motion in transverse directions and in both cases some RR modes, which do indeed form a vector in the IIB case and a 2-form in the IIA case. One checks that the field-strength of the 2-form is self-dual.

In both cases the world-volume effective theory is a 16-supercharge SUSY QFT in 6 dimensions. Some basic facts about 6d 16-SUSY QFTs are reviewed in Box 13.3. As explained there, we have two inequivalent 6d SUSY algebras, types $(1, 1)$ and $(2, 0)$, whose R-symmetries are, respectively

$$(1, 1) \quad Sp(1) \times Sp(1) \simeq SO(4) \quad (13.97)$$

$$(2, 0) \quad Sp(2) \simeq SO(5) \quad (13.98)$$

and a unique matter³³ supermultiplet for each SUSY algebra whose scalar fields form a vector of the R-symmetry group $SO(4)$ and, respectively, $SO(5)$. The scalars living on the 5-brane have the natural interpretation of coordinates for the motion in 4,

³² See the discussion in Chap. 12.

³³ A *matter* supermultiplet is a representation of SUSY which contains only fields of spin ≤ 1 .

BOX 13.3 Facts about 6d supersymmetric QFT

In 6d there are *two* SUSY algebras with 16 SUSYs, types (1, 1) and (2, 0), whose supercharges are, respectively, *two* symplectic-Majorana–Weyl spinors of *opposite* and *equal* chirality. On group theory grounds, the automorphism group (“R-symmetry”) of the 6d SUSY algebra of type (p, q) is [13] $Sp(p) \times Sp(q)$ (other notation: $USp(2p) \times USp(2q)$). Each SUSY algebra has a *unique* matter supermultiplet (i.e. with spins ≤ 1)

non-chiral (1, 1):	vector multiplet A_μ, ϕ^a ($a = 1, \dots, 4$) + fermions
chiral (2, 0):	tensor multiplet $B_{\mu\nu}^+, \chi^a$ ($a = 1, \dots, 5$) + fermions

where the notation $B_{\mu\nu}^+$ means that the 3-form field-strength of the 2-form $B_{\mu\nu}^+$ is constrained to be self-dual. The scalars ϕ^a, χ^a transform as a vector under the automorphism group of the SUSY algebra automorphism group which are respectively

$$Sp(1) \times Sp(1) \simeq SO(4) \quad \text{and} \quad Sp(2) \simeq SO(5)$$

. The R-symmetry groups have the interpretation of rotational symmetries in, respectively, 4 and 5 transverse directions, and the massless scalars ϕ^a (resp. χ^a) are the Goldstone bosons of 4 (resp. 5) broken translations, so they naturally describe the motion of the 5-brane in 10, respectively 11 dimensions. They match the expectations for the Type IIB NS5 and the M-theory M5 brane, respectively.

Since IIA and IIB compactified on a circle are equivalent, this should be also true for these two 6d QFTs, which both should reduce to the 16-SUSY vector multiplet in 5d. In 5d the bosonic fields of the vector multiplet are one vector $A_\mu|_5$ and five scalars Φ^a . Reducing the two 6d supermultiplets one gets the bosonic fields

$$\begin{array}{ll} \text{vector} & A_\mu|_5 = A_\mu, \quad \Phi^5 = A_5, \quad \Phi^a = \phi^a \quad \text{for } a = 1, \dots, 4 \\ \text{vector} & A_\mu|_5 = B_{\mu 5}^+, \quad \Phi^a = \chi^a \quad \text{for } a = 1, \dots, 5 \end{array} \quad (\S)$$

Note that, because of the self-duality constraint, $B_{\mu\nu}^+$ represents the same degrees of freedom of $B_{\mu 5}^+$ (in 5d a 2-form and a vector are dual hence equivalent).

respectively 5 transverse directions, with world-volume R-symmetry identified with spacetime rotations in the orthogonal directions. Thus the world-volume theory with (1, 1) SUSY (resp. (2, 0) SUSY) naturally describe the motion of a dynamical 5-brane in 10 (resp. 11) dimensions. This matches nicely with the physical expectation that the IIB NS5-brane moves in a 10-dimensional space-time, while the IIA NS5, which we identify with the M5-brane of M-theory, moves in an 11-dimensional space.

Unbroken SUSYs Let us check directly that, indeed, the unbroken supersymmetries of a NS5-brane form the (2, 0) SUSY algebra for IIA and the (1, 1) one for IIB. It is a deep fact that the non-chiral string theory IIA has a chiral 5-brane while the chiral IIB theory has a non-chiral 5-brane. The supersymmetry variations of the gravitino fields in a general background are

$$\delta\psi_M = D_M^-\epsilon, \quad \delta\tilde{\psi}_M = D^+\tilde{\epsilon}. \quad (13.99)$$

Here $D_M^\pm$ are the covariant derivative with the spin connection ω replaced by

$$\omega^\pm = \omega \pm \frac{1}{2}H \quad (13.100)$$

with H the NS-NS 3-form field strength. The difference of sign on the two sides occurs because H is odd under worldsheet parity. The modification $\omega \rightarrow \omega^\pm$ can be deduced by writing the supersymmetric world-sheet action for the string moving in a metric and 2-form background [56–58]. The simplest way to get (13.100) is to realize that the 2d world-sheet Lagrangian $\mathcal{L}_{2d}$

$$\mathcal{L}_{2d} = \cdots + i(g_{\mu\nu} + B_{\mu\nu})\psi^\mu \bar{D}\psi^\nu + i(g_{\mu\nu} - B_{\mu\nu})\tilde{\psi}^\mu D\tilde{\psi}^\nu + \cdots \quad (13.101)$$

the left/right movers see an “effective metric”

$$G_{\mu\nu}^\pm = g_{\mu\nu} \pm B_{\mu\nu} \quad (13.102)$$

and “effective Kristoffel symbols”

$$\begin{aligned} \Gamma_{\rho\mu\nu}^\pm &= \frac{1}{2}(\partial_\mu G_{\rho\nu}^\pm + \partial_\nu G_{\mu\rho}^\pm - \partial_\rho G_{\mu\nu}^\pm) = \\ &= \Gamma_{\rho\mu\nu} \pm \frac{1}{2}(\partial_\mu B_{\rho\nu} + \partial_\nu B_{\mu\rho} - \partial_\rho B_{\mu\nu}) = \Gamma_{\rho\mu\nu} \mp \frac{1}{2}H_{\rho\mu\nu}. \end{aligned} \quad (13.103)$$

Under the decomposition of the 10d Lorentz group in the world-volume Lorentz symmetry times the transverse rotations

$$SO(9, 1) \rightarrow SO(5, 1) \times SO(4) \quad (13.104)$$

the 10d spinors decompose as

$$\mathbf{16}_s \rightarrow (\mathbf{4}_s, \mathbf{2}_s) \oplus (\mathbf{4}_c, \mathbf{2}_c), \quad \mathbf{16}_c \rightarrow (\mathbf{4}_s, \mathbf{2}_c) \oplus (\mathbf{4}_c, \mathbf{2}_s). \quad (13.105)$$

The non-zero components of the connection $\omega^\pm$ for the 5-brane solution lie in the transverse Lorentz group

$$SO(4) \simeq SU(2) \times SU(2) \quad (13.106)$$

and for the NS5-branes ω^+ and ω^- have the property that they lie entirely in the first or second $SU(2)$, respectively. Indeed, a background is supersymmetric with 16 supercharges iff the corresponding world-sheet theory is superconformal with (4, 4) enhanced supersymmetry; then there are conserved $SU(2)$ left and right current algebras, and the corresponding $SU(2)$ symmetries should commute with the connections seen (respectively) by the left- and right-movers. We get the same conclusion from

the structure of the spacetime supergravity. A constant spinor charged under the second $SU(2)$, i.e. a $\mathbf{2}_c$ of $SO(4)$, is then annihilated by D_M^+ and one charged under the first $SU(2)$ (a $\mathbf{2}_s$ of $SO(4)$) by D_M^- . These parallel spinors correspond to the unbroken SUSYs.

Thus the *left-moving* SUSYs transforming as a $\mathbf{2}_s$ of $SO(4)$ are unbroken—these are a $\mathbf{4}_s$ of $SO(5, 1)$ both in IIA and IIB. Also unbroken are the *right-moving* SUSYs transforming as a $\mathbf{2}_c$ of $SO(4)$. From Eq. (13.105), we see that they are a $\mathbf{4}_s$ in type IIA (from the $\mathbf{16}_c$) and a $\mathbf{4}_c$ in IIB (from the $\mathbf{16}_s$). In other words—in terms of SUSY on the 6d brane world-volume—for IIA we have 6d (2,0) supersymmetry while for IIB we have 6d (1,1) SUSY as was to be shown.

Note 13.8 We give an alternative argument. The matter supermultiplet of 6d (1, 1) SUSY (resp. 6d (2, 0) SUSY) is a vector multiplet, whose gauge field is a 1-form A_μ (resp. 2-form) whose electric source is a point-particle (resp. a string). By spectral completeness, the 6d QFT should contain charged point-particles (resp. charged strings) which can only be the endpoints of a string (resp. the boundaries of membranes). Since a D1 can terminate on a IIB NS5 we deduce that the IIB NS5 world-volume QFT has (1, 1) SUSY, whereas a D2 can end on a IIA NS5, so the IIA NS5 world-volume QFT should have (2, 0) SUSY.

The Spacetime Detected by the 5-Branes The obvious interpretation of the IIA NS5-brane is as an M-theory M5 which is transverse to the 11th dimension. As in the discussion of the 2-brane, the NS5 should have an additional collective coordinate for its motion in the 11th direction. Again, the 6d QFT on the brane should be smart enough to detect all eleven and not just ten spacetime dimensions. This is sharp contrast to the IIB NS5 whose world-volume QFT sees a spacetime with only ten dimensions.

Indeed the world-volume QFT has the required “magic” properties in both Type II theories. The tensor multiplet of 6d (2, 0) SUSY has five scalars, as many as the transverse directions in M-theory, $5 \equiv 11 - 6$. Four of these scalars are from the NS-NS sector and are collective coordinates for the directions perpendicular to the 5-brane which are visible in string perturbation theory. The fifth scalar, from the RR sector, must be the collective coordinate for the 11-th dimension. Note that the five scalars transform in the vector representation of the $SO(5) \simeq Sp(2)$ R-symmetry of (2, 0) SUSY, so rotational and Poincaré invariances are restored in all transverse directions in 11d. It is remarkable that the IIA 2-brane and 5-brane world-volume theories already know that they secretly live in 11 dimensions.

Note 13.9 We have shown in Eqs. (13.90)–(13.95) that the 3d QFT living on the D2 implies that the 11th dimension is periodic (a circle). The same should be true for the 6d QFT living on the IIA NS5. That this is indeed the case can be shown as follows. Compactify the 6d (2, 0) QFT on a circle S^1 . Since IIA and IIB are T -dual, we should get the *same* 5d QFT by compactifying the IIB NS5 on the T -dual circle $S^{1'}$. From Eq. (S) in Box 13.3 we see that the second 5d theory has 4 non-compact scalars ϕ^a coming from the 6d scalars (the coordinates for the motion of the IIB

NS5 in 10d) and a 5th scalar A_4 which—properly speaking—is the gauge holonomy along the dual circle S^1 ,

$$U = P \exp \oint_{S^1} A \quad (13.107)$$

which takes value in the gauge group $U(1) \simeq S^1$ and so is a *compact* scalar. These 5 scalars, 4 non-compact and one compact, should match the 5 scalars we get by dimensional reduction of the IIA NS5 world-volume theory, which are the same 5 scalars we had originally in the 6d (2, 0) theory. We conclude that the 6d (2, 0) QFT on the IIA NS5 has 1 compact and 4 non-compact scalars. Hence the IIA NS5 world-volume QFT sees a $\mathbb{R}^{9,1} \times S^1$ spacetime:

Type IIA D2 and NS5 world-volume QFTs detect an 11d space-time which is a S^1 -bundle over the 10d perturbatively visible spacetime

These magical properties of the 3d and 6d QFTs living on the M2 and M5 branes yield very convincing evidence for the proposed duality of strongly coupled Type IIA with 11d M-theory.

Further Checks The tension of the IIA NS5 is equal to the one of the IIB NS5

$$\tau_{\text{NS5}} = \frac{1}{(2\pi)^5 g^2 \alpha'^3} = \frac{\tau_{\text{D2}}^2}{2\pi} = \frac{M_{11}^6}{2\pi}. \quad (13.108)$$

Like the tension of the D2, τ_{NS5} is independent of R_{10} as it must be

$$\tau_{\text{D2}} = \tau_{\text{M2}}, \quad \tau_{\text{NS5}} = \tau_{\text{M5}}. \quad (13.109)$$

This also fits with the interpretation of the D4-brane as a wrapped M5

$$\tau_{\text{D4}} = 2\pi R_{10} \tau_{\text{M5}}. \quad (13.110)$$

M2's Ending on M5's: Tensionless Strings Since the IIA NS5 and D2 branes are both localized in the 11-th dimension, the configuration of a D2 ending on a NS5 lifts to an M-theory configuration of an M2 ending on a M5. It is interesting to consider two nearby 5-branes with a 2-brane stretched between them, either in the IIA or M-theory context. The 2-brane is still extended in one-direction and so behaves as a string. Its effective tension is proportional to the distance r between the two 5-branes

$$\tau_1 = r \tau_{\text{M2}}. \quad (13.111)$$

Let us compare this configuration with two nearby (parallel) NS5 branes in IIB with a 1-brane suspended between them. From the viewpoint of the 6d QFT, the

1-brane effectively look like as a vector supermultiplet of particles of mass $r\tau_1$. In the limit $r \rightarrow 0$ the vector particle becomes massless: the unbroken gauge symmetry enhances from $U(1)^2$ to $U(2)$. This is just the Higgs phenomenon: r is the v.e.v. of a Higgs field in the adjoint representation, and when it goes to zero the full non-Abelian symmetry gets restored.

In Type IIA the same argument gives *tensionless strings* as $r \rightarrow 0$. For small r the lightest scale in the theory is set by the tension of these strings. They are totally different from the strings we have studied: they live in six dimensions, are not associated with gravity, and have no adjustable coupling constant—the interactions are of order 1. The world-volume theory on a stack of M5-branes are strongly-interacting 6d (2, 0) SCFT. The existence of conformal field theories in 6d is quite unexpected since no conformal theory described by a Lagrangian may exist in more than 4 dimensions. These SCFTs are inherently strongly coupled, and their existence is a fundamental prediction of string theory. Their mere existence has astonishing implications in many areas of physics, and has greatly improved our knowledge of QFT. We shall say some more words on them in Sect. 14.5.

IIA 6-branes

The D6 is the magnetic source of the RR gauge field $A^{(1)}$ whose electric source is the D0 particle. Since the D0-brane carries Kaluza–Klein (KK) electric charge, the D6 must be a *KK monopole*. We described KK monopoles in Sect. 6.1: geometrically they are the total space of $U(1)$ principal bundles over a lower-dimensional effective spacetime with a non-trivial first Chern class. The D6 is a KK monopole with a 7d Poincaré invariance, which is an extremal solution to the IIA SUGRA e.o.m., in fact a half-BPS soliton. Thus its 11d geometry has the form

$$\mathbb{R}^{6,1} \times M_4 \rightarrow \mathbb{R}^{6,1} \times X_3, \quad (13.112)$$

where X_3 is a non-compact real 3-fold, and $M_4 \rightarrow \mathbb{R}^{6,1}$ a $U(1)$ principal bundle. The non-compact effective 10d space is then $\mathbb{R}^{6,1} \times X_3$. The S^1 fiber is identified with the M-theory circle ($\equiv$ KK circle). As reviewed in Sect. 6.1, a metric on M_4 , invariant by translations along the fibers, defines a connection form θ on this principal bundle,³⁴ which is the (gauge orbit of) KK gauge field. The field strength is the 2-form on $\mathbb{R}^3$

$$F = d\theta. \quad (13.113)$$

The 11d geometry (13.112) is a KK monopole precisely iff the flux of F in $\mathbb{R}^3$ is non-zero, i.e. if the first Chern class of the S^1 -bundle is non-zero.

The metric on the total space M_4 must be such that the full configuration is half-BPS. By the same argument as in Eqs. (13.104)–(13.106), this requires that the Riemann curvature of the 4-fold M_4 is self-dual (or anti-selfdual), i.e. that M_4 is a 4d *gravitational instanton* or, equivalently, a hyperKähler 4-fold.

³⁴ Contrary to the math convention, we take θ to be real, i.e. it is $-i$ times the mathematical definition of connection form.

4d gravitational instantons with the required properties exist: they are known as (generalized) Taub-NUT spaces [59–62]. We defer a detailed discussion of their elegant hyperKähler geometry to Sect. 14.1.

IIA D8-Branes

The M-theory meaning of 8-branes is rather subtle, and best understood from the relation of M-theory with the heterotic $E_8 \times E_8$ string. For this reason, we defer this topic after the discussion of heterotic $E_8 \times E_8$ at strong coupling.

Note 13.10 We should expect the IIA D8 to have a subtle interpretation in M-theory. The D8 is the electric source of the non-dynamical RR 9-form $C^{(9)}$ whose field-strength $F^{10} = dC^{(9)}$ is constant in space-time by effect of its e.o.m.³⁵ We saw in Sect. 8.4 that a non-zero F^{10} background leads to an effective theory which is *massive* IIA SUGRA which, as we remarked in Chap. 8, does *not* lift to 11 dimensions.

13.7 The $E_8 \times E_8$ Heterotic String at Strong Coupling

The last 10d SUSY string theory is the $E_8 \times E_8$ heterotic string. It is T -dual to the $SO(32)$ heterotic string whose strong coupling limit we already know, so we can determine the $g \rightarrow \infty$ limit of the $E_8 \times E_8$ theory by a chain of already established dualities. Our goal in this section is to find a dual weakly coupled description of the 10d $E_8 \times E_8$ string for $g \gg 1$. We need to keep track of how the dilaton Φ and the relevant components of the metric transform at each step along the duality chain.

In each string theory the natural spacetime metric is the one which appears in the F1 world-sheet action. The dualities interchange F1 with other kinds of strings, and the stringy metric in the various descriptions differs because different probes see different spacetime geometries. The chain of dualities produces a map between the initial dilaton and metric to the final ones. We wish the final dilaton to get small when the original one gets large, so producing a weakly coupled dual of the $E_8 \times E_8$ model at strong coupling. In addition the radii of the circles should get big, to avoid light winding states. Since we are interested in understanding the $E_8 \times E_8$ heterotic string moving in $\mathbb{R}^{9,1}$ at large (but finite) coupling, the relevant order of limits is first sending the radii to infinity and only after making the coupling large.

Now we go through each steps of our duality chain:

T -Duality: Heterotic $E_8 \times E_8$ on S^1 to Heterotic $SO(32)$ on S^1 We compactify the heterotic $E_8 \times E_8$ theory with coupling g on a circle of large radius R_9 , and turn on the Wilson line which breaks the gauge symmetry (cf. Sect. 7.7.1)

$$E_8 \times E_8 \rightarrow SO(16) \times SO(16). \quad (13.114)$$

³⁵ Another way of seeing that the gauge potential $C^{(9)}$ does not describe propagating d.o.f. is to notice that the D8 has no dual magnetic source by degree/dimension considerations.

We will eventually take $R_9 \rightarrow \infty$ to get back the 10d theory; in this limit the Wilson line becomes irrelevant. As discussed Sect. 7.7.1, this theory is T -dual to the $SO(32)$ heterotic string again with a Wilson line breaking the gauge group $SO(32) \rightarrow SO(16) \times SO(16)$. The couplings and radii are related as

$$R'_9 \propto R_9^{-1}, \quad g' \propto g R_9^{-1}, \quad (13.115)$$

where primed quantities are for the $SO(32)$ theory and unprimed for the original $E_8 \times E_8$ one. We keep track of the field dependence of each side neglecting $O(1)$ numerical constants

$$R_9 \propto G_{99}^{1/2}, \quad g \propto e^\Phi. \quad (13.116)$$

The transformation of g (13.115) follows from the T -duality Buscher rule; cf. Sect. 6.4.2.

S-Duality: Heterotic $SO(32)$ to Type I on S^1 Now use Type I/heterotic duality to rewrite the above configuration in Type I theory with

$$g_I \propto g'^{-1} \propto g^{-1} R_9, \quad R_{9I} \propto g'^{-1/2} R'_9 \propto g^{-1/2} R_9^{-1/2}. \quad (13.117)$$

The transformation of G_{99} follows from Eq. (13.56).³⁶ We are interested in the limit in which g and R_9 are both large. We can make g_I small by an appropriate order of limits, but the radius of the Type I theory gets small, and we must go to the T -dual description to get a convenient physical picture.

T -Duality: Type I on S^1 to Type IIA on $S^1/\mathbb{Z}_2$ Consider a T -duality in the 9-direction of the Type I theory. The compact dimension becomes a segment of length $\pi\alpha'/R_{9I}$ with eight D8-branes and an orientifold at each end with

$$g_{I'} \propto g_I R_{9I}^{-1} \propto g^{-1/2} R_9^{3/2}, \quad R_{9I'} \propto R_{9I}^{-1} \propto g^{1/2} R_9^{1/2}. \quad (13.118)$$

If we take $g \rightarrow \infty$ at fixed R_9 then we have reached a good description. However, our real interest is the ten-dimensional $E_8 \times E_8$ theory at fixed large coupling, i.e. the proper order of limits is first $R_9 \rightarrow \infty$ and then $g \rightarrow \infty$. The coupling $g_{I'}$ then becomes large, but one final duality will bring us to a good description. The theory that we have reached is often called the Type I' theory.

In the bulk, between the two orientifold planes, we have the IIA theory, so we can also think of Type I' as the IIA theory on the segment $S^1/\mathbb{Z}_2$. The coset must be an orientifold because the only spacetime parity symmetry of the IIA theory includes also a world-sheet parity transformation.³⁷

³⁶ Recall from Sect. 8.7 that $G_{I\mu\nu} = e^{-\Phi_h} G_{h\mu\nu}$ and $\Phi_I = -\Phi_h$.

³⁷ Space-time parity interchanges the gravitini of chirality +1 and chirality -1. The corresponding supercharges are described by left- and right-moving 2d currents, respectively—so spacetime parity symmetry requires a flip in the orientation of the string world-sheet.

S-Duality: Type IIA on $S^1/\mathbb{Z}_2$ to M-theory on $S^1/\mathbb{Z}_2$ As $R_9 \rightarrow \infty$ the IIA theory gets strongly coupled, so the physics between the orientifold planes is described in terms of M-theory with a new periodic dimension. The correct transformation of the fields is obtained from the dimensional reduction of 11d supergravity to 10d and then the field redefinition to the string frame metric, see Eq. (8.44).³⁸ We get

$$R_{10M} \propto g_{I'}^{2/3} \propto g^{-1/3} R_9, \quad R_{9M} \propto g_{I'}^{-1/3} R_{9I'} \propto g^{2/3}. \quad (13.119)$$

As the original R_9 is taken to infinity, the new R_{10M} diverges linearly. Evidently we should identify the original 9-direction with the final 10-direction. Hence at the last step we also renumber the coordinates

$$(9, 10) \longleftrightarrow (10', 9'). \quad (13.120)$$

The final dual for the strongly coupled $E_8 \times E_8$ theory in 10 dimension is M-theory with 10 noncompact dimensions and the $10'$ -direction, which is a transform of the original 9-direction, compactified. This is the same theory as in the strong coupled IIA string with the difference that *now the $10'$ -direction is not a circle but a segment with boundaries at the two orientifold planes.*

In Conclusion: M-theory on S^1 is strongly coupled IIA theory while M-theory on $S^1/\mathbb{Z}_2$ is the strongly coupled $E_8 \times E_8$ heterotic theory. At each end of the segment there is an orientifold plane and eight D8-branes, but now both are nine-dimensional as they bound a 10-dimensional space. The $E_8 \times E_8$ gauge degrees of freedom thus live in these walls, one E_8 in each “end of the world” wall [63].

We summarize the sequence of dualities:

$$\text{het } E_8 \times E_8 \xrightarrow{T_9} \text{het } SO(32) \xrightarrow{S} \text{type } I \xrightarrow{T_9} \text{type } I' \xrightarrow{S} \text{M-theory on } S^1/\mathbb{Z}_2. \quad (13.121)$$

A heterotic string running in the 8-direction becomes

$$F_8 \xrightarrow{T_9} F_8 \xrightarrow{S} D_8 \xrightarrow{T_9} D_{89} \xrightarrow{S} M_{8,10'}, \quad (13.122)$$

that is, a membrane running between the two boundaries, see Fig. 13.1.

The story is highly constrained by anomalies, and this is in fact how it was originally discovered. The $d = 11$ supergravity theory in a space with boundaries has anomalies unless the boundaries carry precisely the E_8 degrees of freedom [63].

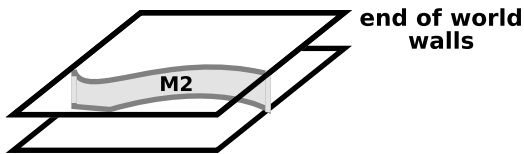
Note also the sequence of dualities

$$p_9 \xrightarrow{T_9} F_9 \xrightarrow{S} D_9 \xrightarrow{T_9} D_\emptyset \xrightarrow{S} p_{10} = p_{9'}, \quad (13.123)$$

which confirms the identification of the original 9-direction with the final 10-direction.

³⁸ Recall from Sect. 8.4 that $G_{\mu\nu}|_{\text{old}} = e^{-\sigma} G_{\mu\nu}|_{\text{new}}$ and $\sigma = \frac{2\Phi}{3}$.

Fig. 13.1 The $E_8 \times E_8$ heterotic string as a M2-brane suspended between the two “end of the world” walls.



13.8 IIA D8-Branes Versus M-Theory

We go back to the issue of D8-branes from the M-theory perspective.

In string theory the D8 is a source for the dilaton. To the linear order, a D8 produces a constant gradient for the dilaton (since the D8 has codimension 1), but the full non-linear supergravity equations for the dilaton, metric, and RR 9-form imply that the dilaton diverges at finite distance from the D8-brane. We recall from Sect. 8.8 the relevant formulae for the SUGRA fields sourced by a Dp -brane

$$ds^2 = H_p^{-1/2} \eta_{\mu\nu} dx^\mu dx^\nu + H_p^{1/2} dy^i dy^i, \quad (13.124)$$

$$e^{2\Phi} = g_s^2 H_p^{(3-p)/2}, \quad (13.125)$$

$$C^{(p+1)} = g_s^{-1} (1 - H_p^{-1}) dx^0 \wedge \cdots dx^p, \quad (13.126)$$

where $\mu = 0, 1, \dots, p, i = p+1, \dots, 9$ and H_p is a harmonic function of the transverse coordinates y^i . A single Dp brane at $r \equiv \sqrt{y^i y^i} = 0$ is given by

$$H_p = 1 + \left(\frac{r_p}{r} \right)^{7-p}, \quad (13.127)$$

where

$$r_p^{7-p} = (2\pi)^{p-2} g_s 2^{7-2p} \pi^{(9-3p)/2} \Gamma\left(\frac{7-p}{2}\right) \alpha'^{(7-p)/2}. \quad (13.128)$$

Since $\Gamma(-1/2) = -2\sqrt{\pi}$, one has

$$r_8^{-1} = g_s 2^{-3} \pi^{-3/2} \Gamma(-1/2) (\alpha')^{-1/2} \equiv -g_s \frac{(\alpha')^{-1/2}}{4\pi} \equiv -|r_8|^{-1} \quad (13.129)$$

which is *negative* and

$$e^{2\Phi} = \frac{g_s^2 |r_8|^{5/2}}{(|r_8| - r)^{5/2}}, \quad (13.130)$$

where now g_s is the value of e^Φ at $r = 0$. We see that the dilaton diverges at a *finite* distance $|r_8|$ from the D8 brane, while the solution does not make sense for larger r . The only way to cure this pathology is to put an “end of the world” (orientifold) wall before the divergence develops. This sets a *maximum distance* between the D8-brane

and the wall, namely

$$\text{max distance} \equiv |r_8| = \frac{4\pi\sqrt{\alpha'}}{g_s}. \quad (13.131)$$

As one goes to the strongly coupled limit $g_s \rightarrow \infty$ the initial value for the dilaton is greater and hence the distance $|r_8|$ is shorter. In the strong coupled limit $g_s \rightarrow \infty$ the D8-branes disappear in the boundary and in the 11d theory there is no way to pull them out. The moduli of their positions in M-theory just become Wilson lines for the E_8 gauge theory living in the boundary. In conclusion: in M-theory there is no independent object corresponding to the IIA D8-branes. This is consistent with (and explains) our observation in Sect. 8.4 that massive IIA SUGRA does not uplift to an 11d theory.

13.9 The Big Picture: *What Is String Theory?*

The picture which emerges from the rich web of dualities between the various SUSY string theories and M-theory may be (morally) summarized by the drawing in Fig. 13.2.

The figure represents (schematically) the *moduli space* of a UNIQUE supersymmetric quantum theory of gravity which has 6 different asymptotic regimes (sometimes called “*corners*” of the moduli space) in which we have a convenient description of the physics which becomes asymptotically correct as we approach the given “corner”.

Five of these asymptotic “corners” are weakly coupled descriptions (with a systematic perturbative expansion) and correspond to the five perturbative superstring theories we have constructed in this textbook, while the sixth one is 11d M-theory meaning a theory living in 11-dimensional spacetime which at low-energy reduced to

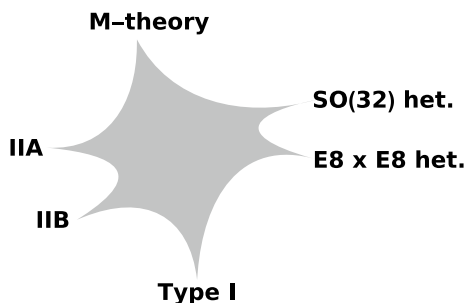


Fig. 13.2 At the non-perturbative level “string theory” is a M(ysterious) theory with a multi-dimensional space of vacua/quantum phases. In the corners of this space some degree of freedom gets weakly-coupled, and we recover asymptotically the various perturbative formulation. In the figure, “M-theory” stands for a regime where 11d SUGRA yields a reliable description.

11d supergravity. By extension, the name “M-theory” is also used for the underlying unified (and Mysterious) theory in the middle of the moduli space.

When a physical question refers to a regime covered by one “corner” we have a systematic way of answering it. We have also some answers for some physical questions in the bulk of the moduli space: this happens when the question refers to a SUSY protected quantity, such as BPS states. Following the BPS states as we go away from the “corner” and go across the moduli space toward a different “corner” was the main tool we used to establish non-perturbative dualities between the corresponding asymptotic descriptions. This procedure also shows that all the nice theories we consider in this book—the five supersymmetric strings and 11d SUGRA—are in fact special limits of a *connected space* of variations of a UNIQUE fundamental theory.

Understanding the physics of this “ultimate” theory which contains Quantum Gravity (and everything else) is the main challenge of Theoretical Physics. In particular we need to learn the basic principles on which is based, and provide an a priori non-perturbative definition of it. An explicit proposal will be presented in Sect. 14.4 of Chap. 14.

References

1. Brennan, T.D., Carta, F., Vafa, C.: The string landscape, the swampland, and the missing corner. [arXiv:1711.00864](#)
2. Weinberg, S.: Ultraviolet divergences in quantum theories of gravitation. In: Hawking, S.W., Israel, W. (eds.) *General Relativity. An Einstein Centenary Survey*. CUP, Cambridge (1979)
3. Lee, S.J., Lerche, W., Weigand, T.: Emergent strings, duality and weak coupling limits for two-form fields. *JHEP* **02**, 096 (2022). [arXiv:1904.06344](#)
4. Lee, S.J., Lerche, W., Weigand, T.: Emergent strings from infinite distance limits. *JHEP* **02**, 190 (2022). [arXiv:1910.01135](#)
5. Witten, E.: String theory dynamics in various dimensions. *Nucl. Phys. B* **443**, 85–126 (1995). [arXiv:hep-th/9503124](#)
6. Hull, C.M., Townsend, P.K.: Unity of superstrings dualities. *Nucl. Phys. B* **438**, 109–137 (1995)
7. Vafa, C.: Lectures on strings and dualities. [arXiv:hep-th/9702201](#)
8. Sen, A.: An introduction to nonperturbative string theory. [arXiv:hep-th/9802051](#)
9. Olive, D., West, P.C.: *Duality and Supersymmetric Theories*. Cambridge University Press, Cambridge (1999)
10. Schwarz, J.H.: An $SL(2, Z)$ multiplet of type IIB superstrings. *Phys. Lett. B* **360**, 13–18 (1995) [erratum: *Phys. Lett. B* **364**, 252 (1995)]. [arXiv:hep-th/9508143](#)
11. Schwarz, J.H.: Superstring dualities. *Nucl. Phys. B Proc. Suppl.* **49**, 183–190 (1996). [arXiv:hep-th/9509148](#)
12. Borel, A., Ji, L.: *Compactifications of Symmetric and Locally Symmetric Spaces*. Birkhäuser, Basel (2005)
13. Cecotti, S.: *Supersymmetric Field Theories. Geometric Structures and Dualities*. Cambridge University Press, Cambridge (2015)
14. Vafa, C.: Evidence for F-theory. *Nucl. Phys. B* **469**, 403–418 (1996). [arXiv:hep-th/9602022](#)
15. Montonen, C., Olive, D.I.: Magnetic monopoles as gauge particles? *Phys. Lett. B* **72**, 117–120 (1977)

16. Coleman, S.R.: The magnetic monopole fifty years later. In: *Proceedings Les Houches Summer School in Theoretical Physics: Gauge Theories in High Energy Physics*, Les Houches, 1981, pp. 461–552 (1982)
17. 't Hooft, G.: Magnetic monopoles in unified gauge theories. *Nucl. Phys.* **B79** 276–284 (1974)
18. Polyakov, A.M.: Particle spectrum in the quantum field theory. *JETP Lett.* **20**, 194–195 (1974)
19. Besse, A.: *Einstein Manifolds*. Springer, Berlin (1987)
20. Helgason, S.: *Differential Geometry, Lie Groups and Symmetric Spaces*. Academic, Cambridge (1978)
21. Borel, A.: *Introduction aux groupes arithmétiques*. Hermann, Paris (1969)
22. Soulé, C.: An introduction to arithmetic groups. In: *Lectures at the Conference Frontiers in Number Theory, Physics and Geometry*, Les Houches, March 9–21, 2003. [arXiv:math/0403390](https://arxiv.org/abs/math/0403390)
23. Morris, D.W.: *Introduction to Arithmetic Groups*. [arXiv:math/0106063](https://arxiv.org/abs/math/0106063)
24. Julia, B.: Group disintegration. In: Hawking, S.W., Roček, M. (eds.) *Superspace and Supergravity*. Cambridge University Press, Cambridge (1981)
25. Chevalley, C.: Certains schémas de groupes semi-simples. *Semin. Bourbaki* **13**, 219 (1961)
26. Steinberg, R.: *Lectures on Chevalley Groups*. Yale University, New Haven. Mimeographed notes (reprinted by the AMS). Available at <https://pdfs.semanticscholar.org>
27. Humphreys, J.E.: *Introduction to Lie Algebras and Representation Theory*. Graduate Texts in Mathematics, vol. 9. Springer, Berlin (1972)
28. Demazure, M.: Schémas en groupes réductifs. *Bull. Math. France* **93**, 369–413 (1965)
29. Demazure, M., Grothendieck, A.: Schémas en groupes. *SGA* **3** (2011)
30. Borel, A.: Density and maximality of arithmetic subgroups. *J. Reine Angew. Math.* **244**, 78–89 (1966)
31. Banks, T., Seiberg, N.: Symmetries and strings in field theory and gravity. *Phys. Rev. D* **83**, 084019 (2011). [arXiv:1011.5120](https://arxiv.org/abs/1011.5120)
32. Satake, I.: *Classification Theory of Semi-Simple Algebraic Groups*. Dekker, New York (1971)
33. Serre, J.-P.: *Complex Semisimple Lie Algebras*. Springer, Berlin (2001)
34. Ooguri, H., Vafa, C.: On the geometry of the string landscape and the swampland. *Nucl. Phys. B* **766**, 21–33 (2007). [arXiv:hep-th/0605264](https://arxiv.org/abs/hep-th/0605264)
35. Vafa, C., Witten, E.: A strong coupling test of *S*-duality. *Nucl. Phys. B* **431**, 3–77 (1994)
36. Göttsche, L.: The Betti numbers of the Hilbert scheme of points on a smooth projective surface. *Math. Ann.* **286**, 193–207 (1990)
37. Bershadsky, M., Vafa, C., Sadov, V.: D strings on D manifolds. *Nucl. Phys. B* **463**, 398–414 (1996). [arXiv:hep-th/9510225](https://arxiv.org/abs/hep-th/9510225)
38. Vafa, C.: Gas of D-branes and Hagedorn density of BPS states. *Nucl. Phys. B* **463**, 415–419 (1996). [arXiv:hep-th/9511088](https://arxiv.org/abs/hep-th/9511088)
39. Vafa, C.: Instantons on D-branes. *Nucl. Phys. B* **463**, 435 (1996). [arXiv:hep-th/9512078](https://arxiv.org/abs/hep-th/9512078)
40. Griffiths, P., Harris, J.: *Principles of Algebraic Geometry*. Wiley, Hoboken (1978)
41. Friedman, R.: *Algebraic Surfaces and Holomorphic Vector Bundles*. Springer, Berlin (1998)
42. Seiberg, N.: Observations on the moduli space of superconformal field theories. *Nucl. Phys. B* **303**, 286–304 (1988)
43. Cecotti, S.: $N = 2$ Landau-Ginzburg versus Calabi-Yau sigma models: nonperturbative aspects. *Int. J. Mod. Phys. A* **6**, 1749 (1991)
44. Huybrechts, D.: *Lectures on K3 Surfaces*. Cambridge University Press, Cambridge (2016)
45. Todorov, A.N.: How many Kähler metrics has a K3 surface. In: Artin, M., Tate, J. (eds.) *Arithmetic and Geometry. Dedicated to I.R. Shafarevitch*. Birkhäuser, Boston (1983)
46. Kobayashi, R., Todorov, A.N.: *Tohoku Math. J.* **39**, 341 (1987)
47. 't Hooft, G.: Computation of the quantum effects due to a four-dimensional pseudoparticle. *Phys. Rev.* **D14** 3432–3450 (1976)
48. Strominger, A.: Heterotic solitons. *Nucl. Phys.* **B343**, 167 (1990)
49. Callan, C.G., Jr., Harvey, J.A., Strominger, A.: World-sheet approach to heterotic instantons and solitons. *Nucl. Phys. B* **359**, 611 (1991)
50. Duff, M.J., Khuri, R.R., Lu, J.X.: String solitons. *Phys. Rep.* **259**, 213–325 (1995). [arXiv:hep-th/9412184](https://arxiv.org/abs/hep-th/9412184)

51. Bröcker, T., Tom Dieck, T.: Representations of Compact Lie Groups. Springer, Berlin (1985)
52. Postnikov, M.M.: Geometry VI. Riemannian Geometry. Encyclopaedia Mathematica Sciences, vol. 91. Springer, Berlin (2001)
53. Fulton, W., Harris, J.: Representation Theory. A First Course. Graduate Texts in Mathematics, vol. 129. Springer, Berlin (1991)
54. Schwarz, J.H.: The power of M theory. Phys. Lett. B **367**, 97–103 (1996). [arXiv:hep-th/9510086](https://arxiv.org/abs/hep-th/9510086)
55. Duff, M.J., Howe, P.S., Inami, T., Stelle, K.S.: Superstrings in $D = 10$ from supermembranes in $D = 11$. Phys. Lett. B **191**, 70 (1987)
56. Hull, C.M., Witten, E.: Supersymmetric sigma models and the heterotic string. Phys. Lett. B **160**, 398–402 (1985)
57. Hull, C.M.: Lectures on Nonlinear Sigma Models and Strings. Cambridge. Print-87-0480.
58. Hull, C.M., Papadopoulos, G., Townsend, P.K.: Potentials for $(p, 0)$ and $(1, 1)$ supersymmetric sigma models with torsion. Phys. Lett. B **316**, 291–297 (1993). [arXiv:hep-th/9307013](https://arxiv.org/abs/hep-th/9307013)
59. Taub, A.H.: Empty space-times admitting a three parameter group of motions. Ann. Math. **53**, 472–490 (1951)
60. Gibbons, G.W., Hawking, S.W.: Gravitational multi-instantons. Phys. Lett. B **78**, 430–432 (1978)
61. Hawking, S.W., Ellis, G.F.R.: The Large Scale Structure of Space-Time. Cambridge Monographs on Mathematical Physics, Cambridge University Press, Cambridge (1973)
62. Anspinwall, P.S., Bridgeland, T., Craw, A., Douglas, M.R., Gross, M., Kapustin, A., Moore, G.W., Segal, G., Szendrői, B., Wilson, P.M.H.: Dirichlet Branes and Mirror Symmetry. Clay Mathematical Monographs, vol. 4. AMS, Providence (2009)
63. Hořava, P., Witten, E.: Heterotic and type I string dynamics from eleven dimensions. Nucl. Phys. B **460**, 506 (1996). [arXiv:hep-th/9510209](https://arxiv.org/abs/hep-th/9510209)

Chapter 14

Applications and Further Topics



Abstract This last chapter contains some topics in non-perturbative superstring theory and basic applications of the duality web between SUSY string theories:

- (a) We discuss the beautiful geometry of Gibbons–Hawking (GH) hyperKähler geometries which describe D6 branes in the M-theory context
- (b) We sketch the mechanism for gauge symmetry enhancement in M-theory
- (c) We complete our discussion of the stringy interpretation of the ADHM construction of YM instantons initiated in Sect. 12.8
- (d) We give a quick introduction to *F-theory*, a far-reaching non-perturbative completion of Type IIB and review its duality with M-theory
- (e) We sketch *Matrix Theory*, a nice proposal for a direct construction of M-theory
- (f) We say a few words about interacting $(2, 0)$ SCFT in six dimensions
- (g) We study the quantum physics of extremal Black Holes in string theory and check that the microstate counting agrees with the Bekenstein–Hawking entropy formula. This fact shows that string theory is really a *consistent* Quantum Gravity.

14.1 Taub-NUT and GH Geometries

We consider a family of Riemannian S^1 fibrations over $\mathbb{R}^3$

$$\pi := M \rightarrow \mathbb{R}^3 \quad (14.1)$$

equipped with Euclidean 4d metrics of the form

$$ds^2 = V(\mathbf{x}) d\mathbf{x} \cdot d\mathbf{x} + V(\mathbf{x})^{-1} (dy + \mathbf{A}(\mathbf{x}) \cdot d\mathbf{x})^2, \quad (14.2)$$

where $\mathbf{x} \equiv (x^1, x^2, x^3)$ are coordinates in $\mathbb{R}^3$ and y is the coordinate along the fiber which is periodically identified (we shall specify its period momentarily). M is then the total space of a $U(1)$ principal bundle and the metric (14.2) is a special instance of the general KK geometry discussed in Sect. 6.1. The circle parametrized by y is the KK circle, $\mathbf{A}(\mathbf{x})$ the KK 3d gauge vector in $\mathbb{R}^3$, and the KK $U(1)$ gauge symmetry is

$$\mathbf{A}(\mathbf{x}) \cdot d\mathbf{x} \rightarrow \mathbf{A}(\mathbf{x}) \cdot d\mathbf{x} + d\Lambda(\mathbf{x}), \quad y \rightarrow y - \Lambda(\mathbf{x}). \quad (14.3)$$

We are interested in Euclidean 4d geometries of the form (14.2) which are *gravitational instantons*, i.e. have a self-dual (or anti-self-dual) Riemann tensor,

$$R^{(2)} = \pm * R^{(2)} \quad \text{or, in the tensor notation, } R_{\mu\nu\rho\sigma} = \pm \frac{1}{2} \epsilon_{\mu\nu\tau\nu} R^{\tau\nu}{}_{\rho\sigma}. \quad (14.4)$$

As we saw in Sect. 11.1.3, an Euclidean 4-manifold is a gravitational instanton (in this sense) iff its metric is hyperKähler. Thus the 4d instantons are, in particular, complex manifolds and holomorphically symplectic, i.e. they have a nowhere vanishing, closed, holomorphic (2, 0)-form Ω .

Claim 14.1 *The metric (14.2) is hyperKähler—i.e. its holonomy belongs to one of the two $SU(2)$'s of $SO(4) \simeq SU(2) \times SU(2)$, equivalently the Riemann tensor is (anti)self-dual—if and only if¹*

$$\nabla V = \pm \nabla \times \mathbf{A}, \quad (14.5)$$

which just says that the function $V := \mathbb{R}^3 \rightarrow \mathbb{R}_{>0}$ is harmonic

$$\Delta V(\mathbf{x}) = 0. \quad (14.6)$$

The metrics (14.2) with $V(\mathbf{x})$ harmonic are called *Gibbons–Hawking (GH) metrics* [1]. We shall show below (Theorem 14.1) that all 4d gravitational instantons having a Killing vector are GH metrics.

Before going to its proof, we state the **claim** a bit more precisely: Eq. (14.6) is allowed to fail at isolated points in $\mathbb{R}^3$ where the S^1 fiber degenerates, i.e. at points where the total space M is *smooth* but the fibration π is not, so the KK gauge field is not well defined there. The function V needs not to be defined at these special points, but its behavior nearby is severely restricted by the condition that the full geometry M extends smoothly over these “bad” points.

In plain English: Eq. (14.6) says that the problem of constructing a GH hyperKähler metric is reduced to solving the Poisson equation of electrostatics with delta-like sources localized at the isolated points where the fiber degenerates

$$\Delta V(\mathbf{x}) = 4\pi \sum_i c_i \delta(\mathbf{x} - \mathbf{x}_i). \quad (14.7)$$

We shall discuss the behavior at the special points $\mathbf{x}_i$ in a moment.

A hyperKähler metric is automatically Ricci-flat², hence a solution of the Einstein equations in vacuum, so a GH metric is a solution to the effective low-energy theory of any superstring model. In fact, it is an *exact* solution to superstring theory since a

¹ The sign $\pm$ reflects a choice of orientation, i.e. self-dual versus anti-self-dual.

² Cf. Sects. 11.1.1 and 11.1.3 or [2].

2d SUSY σ -model with hyperKähler target space is automatically a (4, 4) SCFT, and its β -functions vanish identically by the non-renormalization properties of extended SUSY (see Sect. 11.2).

The KK magnetic field in $\mathbb{R}^3$ is

$$\mathbf{B} \equiv \nabla \times \mathbf{A} = \pm \nabla V, \quad (14.8)$$

and our GH geometry carries a total KK magnetic charge

$$m = \frac{1}{4\pi} \int_{S_\infty^2} \mathbf{B} \cdot d\mathbf{n} = \pm \frac{1}{4\pi} \int_{\mathbb{R}^3} d^3\mathbf{x} \Delta V = \pm \sum_i c_i, \quad (14.9)$$

in other words V is the magnetostatic potential³ of a distribution of point-like monopoles at points $\mathbf{x}_i \in \mathbb{R}^3$ with magnetic charges c_i . Thus

$$V(\mathbf{x}) = \sum_i \frac{c_i}{|\mathbf{x} - \mathbf{x}_i|} + \text{const.} \quad (14.10)$$

The additive constant will be determined by requiring the correct asymptotic geometry as $|\mathbf{x}| \rightarrow \infty$, while the c_i 's are fixed by requiring that the 4d metric is non-singular as $\mathbf{x} \rightarrow \mathbf{x}_i$. Before addressing these two important issues, let us prove Claim 14.1.

Proof Consider the three 2-forms (here we set $\theta \equiv dy + A \equiv dy + \mathbf{A} \cdot d\mathbf{x}$)

$$\omega_1 = dx^1 \wedge \theta + V dx^2 \wedge dx^3 \quad (14.11)$$

$$\omega_2 = dx^2 \wedge \theta + V dx^3 \wedge dx^1 \quad (14.12)$$

$$\omega_3 = dx^3 \wedge \theta + V dx^1 \wedge dx^2. \quad (14.13)$$

Note that

$$\omega_1^2 = \omega_2^2 = \omega_3^2 \text{ is nowhere zero} \quad (14.14)$$

while

$$\omega_a \wedge \omega_b = 0 \text{ for } a \neq b. \quad (14.15)$$

We choose the upper sign in Eq. (14.5) and rewrite it in the form notation

$$*dV = F \equiv d\theta \quad (14.16)$$

where $*$ is the Hodge dual in $\mathbb{R}^3$. Equation (14.16) implies

$$d\omega_a = 0 \text{ for } a = 1, 2, 3. \quad (14.17)$$

Condition (14.16) requires $V(\mathbf{x})$ to be harmonic since $dF = 0$ implies $d * dV = 0$. Conversely, if $V(\mathbf{x})$ is harmonic, the 2-form $F \equiv *dV \in \Omega^2(\mathbb{R}^3)$ is closed and we can find a connection form θ on the S^1 bundle M such that (14.16) holds. Taken together, the conditions (14.11)–(14.17) imply that the 2-forms ω_a are the three Kähler forms of a hyperKähler metric g . Indeed, declaring

$$\omega_1 + i\omega_2 = -(\theta - iVdx^3) \wedge (dx^1 + i dx^2) \quad (14.18)$$

³ A magnetostatic potential behaves exactly as the more usual electrostatic potential by electromagnetic duality. For instance, Eq. (14.10) is just the (magnetic version of the) Coulomb potential.

to be the holomorphic symplectic form Ω , we define a complex structure J on the cotangent space T^*M such that the holomorphic cotangent space is spanned by the 1-forms

$$dx^1 + i dx^2, \quad \theta - i V dx^3, \quad (14.19)$$

so that

$$J(dx^1) = -dx^2, \quad J(dx^3) = -V^{-1}\theta \quad (14.20)$$

$$J(dx^2) = dx^1 \quad J(\theta) = V dx^3. \quad (14.21)$$

The Kähler form with respect to this complex structure is ω_3 by Eqs. (14.14)–(14.15). The relation between the metric g and the Kähler form ω is $g(\zeta, \xi) = \omega(\zeta, J\xi)$, and therefore

$$\begin{aligned} ds^2 &= dx^3 J(\theta) - \theta J(dx^3) + V dx^1 J(dx^2) - V dx^2 J(dx^1) = \\ &= V(dx^3)^2 + V^{-1}\theta^2 + V(dx^1)^2 + V(dx^2)^2 = \\ &= V d\mathbf{x} \cdot d\mathbf{x} + V^{-1}\theta^2, \end{aligned} \quad (14.22)$$

which is our metric (14.2). □

Note 14.1 Equation (14.16) should be compared with the Bogomolny equations for an extremal monopole in $\mathbb{R}^3$ [3, 4] (which describes a hyperholomorphic connection in $\mathbb{R}^4$ invariant under translation in one direction [5]) (See Box 14.1).

Regularity at the Special Points $\mathbf{x}_i$

We write (r, θ, ϕ) for the standard polar coordinates in $\mathbb{R}^3$ and we focus on the GH metric with

$$V = 1 + \frac{c}{r} \Rightarrow A = c \cos \theta d\phi, \quad (14.23)$$

that is,

$$ds^2 = \left(1 + \frac{c}{r}\right) (dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2) + \frac{1}{1 + c/r} (dy + c \cos \theta d\phi)^2. \quad (14.24)$$

In the vicinity of the origin, $r \approx 0$, this metric becomes

$$ds^2 \approx \frac{c}{r} dr^2 + cr \left(d\theta^2 + \sin^2 \theta d\phi^2 + (dy/c + \cos \theta d\phi)^2 \right). \quad (14.25)$$

At the origin the length of the circle goes to zero, and the S^1 fiber degenerates to a point. We want the total geometry M to remain smooth. Comparing with Box 14.2, we see that (14.25) is c times the flat metric on $\mathbb{R}^4$, hence regular as $r \rightarrow 0$, *provided we identify y/c with the angle ψ* . Thus the geometry is regular (smooth and complete) precisely iff y/c is periodic of period 4π , i.e. iff

$$y \sim y + 4\pi c, \quad (14.26)$$

otherwise we have a deficit angle at the origin, i.e. a conical singularity.

BOX 14.1 BPS monopole equations versus hyperholomorphic bundles

We identify classical *static* 4d solutions of a field theory with 3d Euclidean solutions of the dimensional reduced theory. The 3d action is the mass M of the 4d soliton. In this vein, we consider a 3d Yang–Mills theory with gauge group $SO(3)$ coupled to an adjoint Higgs field Φ

$$M = \int \text{tr} \left(F \wedge *F + D\Phi \wedge *D\Phi \right) + \lambda \int d^3x \left(|\Phi|^2 - v^2 \right)^2.$$

The v.e.v. $\langle \Phi \rangle = v$ breaks $SO(3) \rightarrow U(1)$. We are interested in the *extremal limit* $\lambda \rightarrow 0$:

$$M = \int \text{tr} \left[(F \pm *D\Phi) \wedge *(F \pm *D\Phi) \right] \pm 2 \int d \text{tr}(F \Phi)$$

This gives a bound (the original Bogomolny bound!)

$$M \geq \left| \int_{S_\infty^2} \text{tr}(\Phi F) \right| = 2v |Q_m|$$

where Q_m is the $U(1)$ magnetic charge. The bound is saturated iff $F = \mp *D\Phi$. If we take Φ and A in the subalgebra $\mathfrak{u}(1) \subset \mathfrak{so}(3)$ and identify $\Phi \equiv V$ we get (14.16).

Hyperholomorphic Bundles

Recall that a bundle with connection ∇ on a complex manifold X is holomorphic iff its curvature form ∇^2 is of pure type $(1,1)$. Let M be a HK manifold. M is a complex manifold for a $\mathbb{P}^1$ -family of complex structures. For each complex structure $\zeta \in \mathbb{P}^1$, a differential form η decomposes into forms of definite type (p, q) . The type of decomposition depends on the complex structure ζ . Typically a bundle with connection is holomorphic only for some particular choices of the complex structure. However there is a “rare” possibility:

Definition 14.1 A bundle with connection ∇ on a HK manifold M is *hyperholomorphic* iff ∇^2 is of type $(1,1)$ in all complex structures, i.e. the bundle is holomorphic in all complex structures.

Hyperholomorphic Bundles on $\mathbb{R}^3 \times \mathbb{R}$ See $\mathbb{R}^4 \simeq \mathbb{C}^2$ as a flat hyperKähler manifold with holomorphic coordinates in complex structure $\zeta \in \mathbb{P}^1$

$$w_\zeta^1 = z^1 - \zeta \bar{z}^2, \quad w_\zeta^2 = \frac{1}{\zeta} z^2 + \bar{z}^1$$

with the family of symplectic/Kähler forms $dw_\zeta^1 \wedge dw_\zeta^2$. The following three 2-forms have type $(1, 1)$ in all complex structures

$$\begin{aligned} dz^1 \wedge d\bar{z}^2 &= (\zeta + \bar{\zeta}^{-1})^{-1} dw_\zeta^1 \wedge d\bar{w}_\zeta^2, \quad dz^2 \wedge d\bar{z}^1 = (\bar{\zeta} + \zeta^{-1})^{-1} dw_\zeta^2 \wedge d\bar{w}_\zeta^1, \\ dz^1 \wedge d\bar{z}^1 - dz^2 \wedge d\bar{z}^2 &= (1 + \zeta \bar{\zeta})^{-1} (dw_\zeta^1 \wedge d\bar{w}_\zeta^1 - \zeta \bar{\zeta} dw_\zeta^2 \wedge d\bar{w}_\zeta^2), \end{aligned}$$

from which we see that the field strength F is $(1,1)$ in all complex structures iff it is of type $(1,1)$ in the complex structure with holomorphic coordinates (z^1, z^2) and satisfies the condition $(dz^1 \wedge d\bar{z}^1 + dz^2 \wedge d\bar{z}^2) \wedge F = 0$ which by Hodge theorem is the same as F being *anti-self-dual*. Hence on a 4d HK manifold a connection A is hyperholomorphic iff it is an anti-instanton. Reducing to 3d we get $F = - *_3 dA_4$ which is the Bogomolny equation with $A_4 \equiv \Phi$

BOX 14.2 Flat metric in $\mathbb{R}^4$ as a Gibbons–Hawking metric

By Theorem 14.1 all 4d hyperKähler metrics with a $U(1)$ tri-holomorphic symmetry are of the Gibbons–Hawking form. This holds, in particular, for the flat metric. To see this, we first write the round metric on the sphere S^3 in convenient coordinates to emphasize the Hopf fibration with S^1 fibers. We see S^3 as the locus in $\mathbb{C}^2$ with $|z_1|^2 + |z_2|^2 = 1$. Then we parametrize

$$\begin{aligned} z_1 &= e^{i(\phi+\psi)/2} \cos(\theta/2) \\ z_2 &= e^{i(\phi-\psi)/2} \sin(\theta/2), \end{aligned}$$

where ψ is the coordinate along the S^1 fiber. In these coordinates the round S^3 metric reads

$$d\Omega_3^2 = |dz_1|^2 + |dz_2|^2 = \frac{1}{4} (d\theta^2 + d\psi^2 + d\phi^2 + 2\cos(\theta)d\phi d\psi).$$

Setting $R = 2\sqrt{r}$, the flat metric on $\mathbb{R}^4$

$$ds_{\text{flat}}^2 = dR^2 + R^2 d\Omega_3^2$$

becomes

$$\begin{aligned} ds_{\text{flat}}^2 &= \frac{dr^2}{r} + r(d\theta^2 + d\psi^2 + d\phi^2 + 2\cos(\theta)d\phi d\psi) = \\ &= \frac{1}{r} (dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2) + r(d\psi + \cos(\theta)d\phi)^2 \end{aligned}$$

which is the GH metric for the harmonic function $V = 1/r$. Note that $0 \leq \theta \leq \pi$ while ψ and ϕ are periodic of period 4π (but a shift of both angles by 2π is also trivial). Thus the metric above is smooth at $r = 0$ if and only if we identify $\psi \sim \psi + 4\pi$.

Note 14.2 We have defined a fibration $\mathbb{C}^2 \setminus \{0\} \rightarrow \mathbb{R}^3 \setminus \{0\}$ where the coordinates of $\mathbb{C}^2$ are

$$Z_1 = R z_1, \quad Z_2 = R z_2 \quad |z_1|^2 + |z_2|^2 = 1,$$

and those of $\mathbb{R}^3$

$$x_1 + ix_2 = r e^{i\psi} \sin \theta, \quad x_3 = r \cos \theta$$

given by

$$(x_1 + ix_2, x_3) = \left(\frac{1}{2} Z_1 Z_2, \frac{1}{4} (|Z_1|^2 - |Z_2|^2) \right)$$

i.e. by the 3 momentum maps of the rotation in ϕ . Same story with $\psi \leftrightarrow \phi$ and $Z_2 \leftrightarrow \bar{Z}_2$.

Thus regularity correlates the coefficient of the poles of $V(\mathbf{x})$ with the periodicity of y , and hence the constants c_i are equal for all points $\mathbf{x}_i$. In the standard normalization y has period 4π , and hence regularity holds if $V(\mathbf{x})$ has the form

$$V(\mathbf{x}) = v + \sum_{i=1}^N \frac{1}{|\mathbf{x} - \mathbf{x}_i|}, \quad (14.27)$$

for a certain constant v . If $v \neq 0$ the GH metric is known as the *multi-center Taub-NUT metric*. The original Taub-NUT metric is the case $N = 1$ [6]. It represents a KK monopole in $\mathbb{R}^3$ of unit magnetic charge (see Eq. (14.9)).

A_{N-1} Singularities

For comparison, suppose that $V(\mathbf{x})$ is the magnetostatic potential for a point-charge of strength $N \in \mathbb{N}$

$$V(\mathbf{x}) = v + \frac{N}{|\mathbf{x}|}. \quad (14.28)$$

Comparing Eq. (14.25) with the identification $\psi \equiv y/c \equiv y/N$ with Box 14.2 we see that near the origin we have

$$z_1 = e^{i\psi/(2N)} e^{i\phi/2} \cos(\theta/2) \quad z_2 = e^{i\psi/(2N)} e^{-i\phi/2} \cos(\theta/2) \quad (14.29)$$

so that $\psi \sim \psi + 4\pi$ means $(z_1, z_2) \sim e^{2\pi i/N} (z_1, z_2)$, that is, at the origin we have a $\mathbb{C}^2/\mathbb{Z}_N$ singularity, i.e. an A_{N-1} du Val singularity; cf. Appendix 2 to Chap. 2.

ALE Spaces

Taking $v = 0$ we get more general gravitational instantons known as the ALE⁴ spaces. The simplest one been the Eguchi–Hanson gravitational instanton [7]. ALE spaces are related to the McKay correspondence (Appendix 2 to Chap. 2): they are the smooth complex surfaces X_Γ which resolve the canonical singularities $\mathbb{C}^2/\Gamma$ with Γ a finite subgroup of $SU(2)$. The 2-form $dz^1 \wedge dz^2 \in \Omega^2(\mathbb{C}^2)$ is Γ invariant so it lifts to the resolution X_Γ which is then crepant. Therefore X_Γ is a holomorphic symplectic space with a natural hyperKähler metric which is the ALE metric. The simplest case $\Gamma = \mathbb{Z}_2$ is the original Eguchi–Hanson instanton which is just the GH metric with

$$V(\mathbf{x}) = \frac{1}{|\mathbf{x} - \mathbf{x}_1|} + \frac{1}{|\mathbf{x} - \mathbf{x}_2|}, \quad \mathbf{x}_1 \neq \mathbf{x}_2. \quad (14.30)$$

The ALE hyperKähler metrics for all finite subgroups $\Gamma \subset SU(2)$ can be obtained directly using the hyperKähler quotient (cf. Box 12.12). We refer to the literature [8–12] for details.

A General Theorem

A vector field v on a HK manifold is *tri-holomorphic* iff it leaves invariant all Kähler forms, i.e. $\mathcal{L}_v \omega_a = 0$ for $a = 1, 2, 3$. v is then automatically a Killing vector.

Theorem 14.1 *Let M be a hyperKähler 4-manifold with one tri-holomorphic Killing vector K acting freely. Then it is a GH manifold.*

Proof Let ω_a , $a = 1, 2, 3$ be the three orthogonal Kähler forms. By definition, the vector field K leaves invariant the Kähler forms

$$\forall a \quad \mathcal{L}_K \omega_a = 0 \quad \Rightarrow \quad i_K \omega_a = dx_a \quad (14.31)$$

⁴ ALE space= asymptotically locally Euclidean space.

for some functions x_a , called the *3-momentum maps* (cf. Box 12.12). Let y be the coordinate along the orbits of K , so that

$$i_K dy = 1, \quad i_K dx_a = 0. \quad (14.32)$$

Then

$$\omega_a = dy \wedge dx_a + \frac{1}{2} F_{a,bc} dx_b \wedge dx_c \quad (14.33)$$

for some functions $F_{a,bc} = -F_{a,cb}$. From $\omega_a \wedge \omega_b \propto \delta_{ab}$ we get

$$\begin{aligned} dy \wedge dx_a \wedge F_{b,cd} dx_c \wedge dx_d + dy \wedge dx_b \wedge F_{a,cd} dx_c \wedge dx_d = \\ = 4V \delta_{ab} dy \wedge dx_1 \wedge dx_2 \wedge dx_3 \end{aligned} \quad (14.34)$$

for some scalar function V , which implies

$$\frac{1}{2} F_{a,bc} dx_b \wedge dx_c = A_b dx_b \wedge dx_a + \frac{1}{2} V \epsilon_{abc} dx_b \wedge dx_c \quad (14.35)$$

for a certain 1-form $A \equiv A_b dx_b$ and function V which depend on x_a only. Then

$$\omega_a = (dy + A) \wedge dx_a + \frac{1}{2} V \epsilon_{abc} dx_b \wedge dx_c. \quad (14.36)$$

The condition $d\omega_a = 0$ gives

$$dA = *dV. \quad (14.37)$$

Corollary 14.1 *In all GH geometries the S^1 fibration $\pi := M \rightarrow \mathbb{R}^3$ is given by the 3-momentum map of the $U(1)$ Killing vector along the fibers.*

14.1.1 Half-BPS 6-Branes in M-Theory. Non-Abelian Gauge Symmetry

We argued in Sect. 13.6 that the D6 brane of Type IIA gets uplifted in M-theory to a half-BPS soliton of 11d SUGRA with 7d Poincaré invariance which carries one unit of *magnetic* Kaluza–Klein charge. The 11d geometry $\mathbb{R}^{6,1} \times \text{Taub-NUT}$

$$\begin{aligned} ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu + \left(v + \frac{1}{r}\right) (dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2) + \\ + \frac{1}{v + 1/r} (dy + \cos \theta d\phi)^2, \quad v \neq 0 \quad \mu, \nu = 0, 1, \dots, 6, \end{aligned} \quad (14.38)$$

has all these properties. In particular this geometry has 16 parallel spinors (cf. Sect. 11.1.1) so, as a solution to 11d SUGRA, it preserves 16 supercharges, i.e. it is a half-BPS solitonic 6-brane which should be identified with the M-theory uplift of the IIA D6 brane. As $r \rightarrow \infty$ the M-theory circle parametrized by y has radius

$$g_\infty \sqrt{\alpha'} \equiv R_\infty = \frac{2}{v}. \quad (14.39)$$

The constant v then yields the string coupling constant g_∞ far away from the brane.

Note 14.3 The Minkowskian version of the Taub-NUT metric has a mass equal (in appropriate units) to the KK magnetic charge, as it should be for a BPS configuration.

Multi-Center Taub-NUT Geometry and Non-Abelian D.O.F.

A multi-center Taub-NUT geometry with N special points $\mathbf{x}_i$ —whose harmonic function $V(\mathbf{x})$ is as in Eq. (14.27)—is the 11d solution which uplifts to M-theory a configuration of N parallel D6 branes with separations $|\mathbf{x}_i - \mathbf{x}_j|$ in the transverse $\mathbb{R}^3$. The multi-brane system is also a half-BPS solution since there is no net force between the D6's.

From the D6's picture we know that for large separations the light degrees of freedom living on the world-volume are vector supermultiplets in the Cartan subalgebra of $\mathfrak{su}(N)$,⁵ whose scalars' v.e.v. yield the transverse positions $\mathbf{x}_i$ of the branes. For small separations we have in addition light vector supermultiplets associated to the roots $\{e_i - e_j := i \neq j\}$ of $\mathfrak{su}(N)$: as $|\mathbf{x}_i - \mathbf{x}_j| \rightarrow 0$ the full $SU(N)$ gauge symmetry gets restored. From the IIA viewpoint the root vector multiplet associated to the root $e_i - e_j$ are the light states of strings suspended between the i -th and j -th branes.

From the M-theory perspective the non-Abelian d.o.f. associated to the root $e_i - e_j$ have a simple geometric interpretation. Let $\mathbf{x}_i \neq \mathbf{x}_j$ be two special points in $\mathbb{R}^3$ over which the S^1 fiber degenerates to a point and

$$I_{ij} = \{s \mathbf{x}_j + (1-s) \mathbf{x}_i := 0 \leq s \leq 1\} \subset \mathbb{R}^3 \quad (14.40)$$

be the segment connecting them. Its preimage $\pi^{-1}(I_{ij}) \subset M$ in the total space of the S^1 -fibration is topologically a 2-sphere S^2_{ij} : indeed, since the fiber contracts to a point at the endpoints of I_{ij} , $\pi^{-1}(I_{ij})$ is a simply-connected closed surface (see Fig. 14.1). Now consider a M2-brane wrapped on the sphere S^2_{ij} . It has a mass given by the M2 tension times its area

$$\int_0^1 ds |\mathbf{x}_i - \mathbf{x}_j| \sqrt{V(s\mathbf{x}_j + (1-s)\mathbf{x}_i)} \int_0^{4\pi} \frac{dy}{\sqrt{V(s\mathbf{x}_j + (1-s)\mathbf{x}_i)}} = 4\pi |\mathbf{x}_i - \mathbf{x}_j| \quad (14.41)$$

which goes linearly to zero with the separation, as expected from the Type IIA picture where the gauge vectors take mass via the Higgs mechanism.

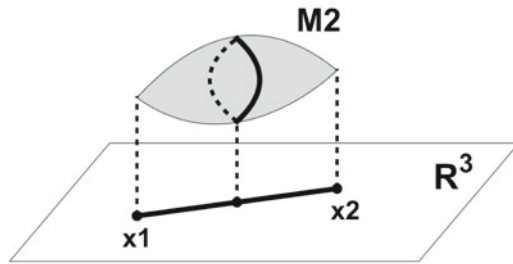
Relation to the McKay Correspondence

In the limit where the N branes are on top of each other we get the GH geometry whose harmonic function $V(\mathbf{x})$ is the electrostatic potential generated by N units of charge at the origin of $\mathbb{R}^3$, Eq. (14.28). As shown above this geometry has a $\mathbb{C}^2/\mathbb{Z}_N$ singularity, that is, a canonical singularity of type A_{N-1} [13, 14]. The *crepant resolution*⁶ of this

⁵ Plus a decoupled free $U(1)$ sector associated to the center-of-mass of the whole system.

⁶ If S is a complex space with an isolated singularity at $p \in S$, by a *resolution* we mean a holomorphic map $f := X \rightarrow S$, with X smooth, which is an isomorphism away from p . The resolution is *crepant* if it preserves the canonical bundle, i.e. $K_X = f^*K_S$.

Fig. 14.1 A M2 brane wrapping the sphere fibered in circles over a segment in $\mathbb{R}^3$ joining the transverse positions $\mathbf{x}_1$ and $\mathbf{x}_2$ of two M-theory unit-charge KK monopoles (corresponding to IIA D6 branes).



singularity is a hyperKähler surface $X_{A_{N-1}} \rightarrow \mathbb{C}^2/\mathbb{Z}_N$ whose fiber over the origin has the topology of a bouquet of $(N - 1)$ 2-spheres (by Milnor theorem [15]) or, in another language, it is a divisor consisting of $(N - 1)$ copies of $\mathbb{P}^1$ whose intersection matrix is given by the adjacency matrix of the A_{N-1} Dynkin graph. This is nothing else than the McKay correspondence which associates to the subgroup $\mathbb{Z}_N \subset SU(2)$ the Lie algebra A_{N-1} (Appendix 2 of Chap. 2). The M2 wrapping these $\mathbb{P}^1$ (of zero volume in the singular limit) yield the massless non-Abelian degrees of freedom of $SU(N)$ SYM.

Generalizing our discussion above for the A_{N-1} singularity, in M-theory a singularity of the form $\mathbb{C}^2/\Gamma$ ($\Gamma \subset SU(2)$ finite) produces 6-branes carrying the d.o.f. of a 16-SUSY SYM with the gauge Lie algebra $\mathfrak{L}_\Gamma$ associated to the finite subgroup $\Gamma \subset SU(2)$ by the McKay correspondence in Appendix 2 of Chap. 2. The Milnor fiber is a configuration of 2-spheres with intersection numbers given by minus the Cartan matrix of the Lie algebra $\mathfrak{L}_\Gamma \in ADE$, and the massless vector multiplet associated to a root of $\mathfrak{L}_\Gamma$ is a M2-brane wrapping the corresponding sphere (cf. Fig. 14.1).

Note 14.4 In addition M-theory contains subtler 6-branes associated to the so-called “frozen” singularities, see [16, 17].

IIA on K3 Revisited

In Sect. 13.3 we sketched the duality between heterotic on T^4 and IIA on K3. Heterotic on T^4 was studied in Sect. 7.7 where we saw that at special points in the moduli spaces the gauge symmetry enhances to a non-Abelian group. The same phenomenon can be described in the language of IIA on K3. Now the special points in the moduli spaces are where the internal K3 develop a singularity, which must be canonical in order not to spoil the hyperKähler property. Locally near such a singularity K3 looks like $\mathbb{C}^2/\Gamma$ for $\Gamma \subset SU(2)$, and the above discussion of gauge symmetry enhancement applies verbatim with M2 replaced by their D2 counterparts. Again the non-Abelian gauge group is given by the McKay correspondence.

14.2 ADHM Construction Versus D-Branes

We have seen in Sect. 12.8 that, along the Higgs branch of the $D(-1)$ “world-volume” theory, a BPS bound state of k $D(-1)$ instantons with N $E3$ ’s ($\equiv$ Euclidean $D3$ ’s) gets identified with a $SU(N)$ instanton of topological charge k of the super-Yang–Mills theory living on the world-volume of the $E3$ ’s. Now we wish to show that the brane construction yields the *precise* gauge field $A_\mu(x)$ of the instanton, so reproducing and explaining the ADHM construction of the general anti-self-dual solution of the Yang–Mills equations, originally obtained by twistorial and Algebro-Geometric methods [18–21].

From the ADHM Data to the Instanton Field Configuration

We saw in Sect. 12.8 that the theory on the $D(-1)$ “world-volume” is a zero-dimensional QFT with gauge group $U(k)$ and flavor group $U(N)$. Its field content consists of two 0d fields (i.e. matrices) B_1, B_2 in the $\mathbf{k} \otimes \bar{\mathbf{k}}$ of $U(k)$ and two 0d fields $Q, \tilde{Q}$ in the $(\mathbf{k}, N)$ of $U(k) \times U(N)$ (cf. Eq. (12.214)). The 0d configuration space $(B_1, B_2, Q, \tilde{Q})$ is then a copy of $\mathbb{C}^{2k(N+k)}$ endowed with the natural flat hyperKähler metric, with a three-holomorphic action of $U(k)$. The 0d e.o.m. are the algebraic equations (12.215) obtained by setting the $U(k)$ hyperKähler momentum map to zero:

$$[B_1, B_2] + Q\tilde{Q}^\dagger = 0, \quad [B_1, B_1^\dagger] + [B_2, B_2^\dagger] + QQ^\dagger - \tilde{Q}\tilde{Q}^\dagger = 0. \quad (14.42)$$

The framed moduli space $\mathcal{M}_{N,k}$ of $SU(N)$ instantons of topological charge k is then the hyperKähler quotient of $\mathbb{C}^{2k(N+k)}$ by this $U(k)$ action (cf. Sect. 12.8, Box 12.12).

It remains to write the explicit gauge field A_μ which corresponds to a ADHM datum, i.e. to a point in $\mathcal{M}_{N,k}$ [18–21]. We identify the Euclidean 4d $E3$ world-volume with $\mathbb{R}^4 \simeq \mathbb{C}^2 \simeq \mathbb{H}$ combining the four coordinates x_i into the quaternion

$$\mathbf{x} = \begin{pmatrix} \bar{z}_2 & -z_1 \\ \bar{z}_1 & z_2 \end{pmatrix} \equiv \begin{pmatrix} x_4 - ix_3 & -x_2 - ix_1 \\ x_2 - ix_1 & x_4 + ix_3 \end{pmatrix}. \quad (14.43)$$

Consider the $2k \times (N + 2k)$ matrix⁷

$$\Delta \equiv \Delta(\mathbf{x}) \stackrel{\text{def}}{=} \begin{pmatrix} Q & B_2 + z_2 & B_1 + z_1 \\ \tilde{Q} & -B_1^\dagger - \bar{z}_1 & B_2^\dagger + \bar{z}_2 \end{pmatrix}, \quad (14.44)$$

which satisfies

$$\partial_\mu \Delta^\dagger \equiv \begin{pmatrix} 0 \\ \partial_\mu \mathbf{x} \end{pmatrix} = \begin{pmatrix} 0 \\ -i\sigma_\mu \otimes \mathbf{1}_k \end{pmatrix} \quad (14.45)$$

where $\sigma_\mu \equiv (\sigma_i, i\mathbf{1}_2)$ are the four 4d Weyl σ -matrices see Box 14.3 for our conventions. The $D(-1)$ e.o.m. (14.42) are equivalent to the matrix factorization

⁷ Here and below we use the block-matrix notation. The 1st block has size N and 2nd block size $2k$.

BOX 14.3 Properties of Euclidean 4d Weyl σ -matrices

We adopt the conventions of [22]. We define ($\mu = 1, 2, 3, 4, i = 1, 2, 3$)

$$\sigma_\mu = (\sigma_i, i\mathbf{1}_2), \quad \bar{\sigma}_\mu = \sigma_\mu^\dagger = (\sigma_i, -i\mathbf{1}_2).$$

The 4d Euclidean Dirac γ -matrices in a chiral basis read (in block-matrix notation)

$$\gamma^\mu = \begin{pmatrix} 0 & -i\sigma^\mu \\ i\bar{\sigma}^\mu & 0 \end{pmatrix}, \quad \gamma^5 \equiv \gamma^1\gamma^2\gamma^3\gamma^4 = \begin{pmatrix} \mathbf{1}_2 & 0 \\ 0 & -\mathbf{1}_2 \end{pmatrix}$$

The generators of $\text{Spin}(4)$ read

$$M_{\mu\nu} \equiv \frac{1}{4}(\gamma_\mu\gamma_\nu - \gamma_\nu\gamma_\mu) = \frac{1}{2} \begin{pmatrix} \sigma^{\mu\nu} & 0 \\ 0 & \bar{\sigma}^{\mu\nu} \end{pmatrix}$$

where

$$\sigma^{\mu\nu} = \frac{1}{2}(\sigma^\mu\bar{\sigma}^\nu - \sigma^\nu\bar{\sigma}^\mu), \quad \bar{\sigma}^{\mu\nu} = \frac{1}{2}(\bar{\sigma}^\mu\sigma^\nu - \bar{\sigma}^\nu\sigma^\mu)$$

$\sigma_{\mu\nu}$ (resp. $\bar{\sigma}_{\mu\nu}$) is anti-self-dual (resp. self-dual) in the 4d indices μ, ν

$$\sigma_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\sigma^{\rho\sigma}, \quad \bar{\sigma}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}\bar{\sigma}^{\rho\sigma}$$

One has

$$\begin{aligned} \frac{1}{2}[\sigma_{\mu\nu}, \sigma_{\rho\sigma}] &= \delta_{\nu\rho}\sigma_{\mu\sigma} - \delta_{\nu\sigma}\sigma_{\mu\rho} - \delta_{\mu\rho}\sigma_{\nu\sigma} + \delta_{\mu\sigma}\sigma_{\nu\rho} \\ \{\sigma_{\mu\nu}, \sigma_{\rho\sigma}\} &= 2(\delta_{\mu\sigma}\delta_{\nu\rho} - \delta_{\mu\rho}\delta_{\nu\sigma}) + 2\epsilon_{\mu\nu\rho\sigma} \end{aligned}$$

$$\Delta\Delta^\dagger = \begin{pmatrix} f^{-1} & 0 \\ 0 & f^{-1} \end{pmatrix} \quad (14.46)$$

where $f \equiv f(\mathbf{x})$ is a $k \times k$ matrix (necessary Hermitian). Then the Hermitian matrix

$$P \equiv P^\dagger = \Delta^\dagger \begin{pmatrix} f & 0 \\ 0 & f \end{pmatrix} \Delta \equiv \Delta^\dagger (\mathbf{1}_2 \otimes f) \Delta \quad P := \mathbb{C}^{N+2k} \rightarrow \mathbb{C}^{N+2k} \quad (14.47)$$

is a projector $P^2 = P$. For generic $\mathbf{x}$ the image of P has dimension $2k$. Consider the orthogonal complementary space, i.e. the kernel of Δ which (generically) has dimension N . We choose an orthonormal basis of $\ker \Delta$ which we assemble into a $(N + 2k) \times N$ matrix $U \equiv U(\mathbf{x})$ (the basis elements are the columns of U). One has

$$P + UU^\dagger = \mathbf{1}_{N+2k}, \quad U^\dagger U = \mathbf{1}_N, \quad \Delta U = 0. \quad (14.48)$$

Claim 14.2 *The connection in $\mathbb{R}^4$*

$$A \equiv A_\mu(\mathbf{x}) dx^\mu \stackrel{\text{def}}{=} U(\mathbf{x})^\dagger dU(\mathbf{x}) \equiv -(dU^\dagger)U \quad (14.49)$$

BOX 14.4 $SU(N)$ instantons over $\mathbb{P}^1(\mathbb{H})$

We extend the $2k \times (N + 2k)$ matrix $\Delta(\mathbf{x})$ in Eq. (14.44), which depends on one quaternion $\mathbf{x}$, to a $2k \times (N + 2k)$ matrix $\Delta(\mathbf{x}, \mathbf{y})$ which depends on *two* quaternions

$$\mathbf{x} = \begin{pmatrix} \bar{z}_2 & -z_1 \\ \bar{z}_1 & z_2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} \bar{w}_2 & -w_1 \\ \bar{w}_1 & w_2 \end{pmatrix}$$

by setting

$$\Delta(\mathbf{x}, \mathbf{y}) = \begin{pmatrix} w_2 Q + w_1 \tilde{Q} & w_2 B_2 - w_1 B_1^\dagger + z_2 & w_2 B_2 - w_1 B_2^\dagger + z_1 \\ -\bar{w}_1 Q + \bar{w}_2 \tilde{Q} & \bar{w}_1 B_2 - \bar{w}_2 B_1^\dagger - \bar{z}_1 & -\bar{w}_1 B_1 + \bar{w}_2 B_2^\dagger + \bar{z}_2 \end{pmatrix}.$$

Note that $\Delta(\mathbf{x}) \equiv \Delta(\mathbf{x}, \mathbf{1})$. For all $q \in \mathbb{H}$

$$\Delta(\mathbf{x}q, \mathbf{y}q) = \bar{q} \Delta(\mathbf{x}, \mathbf{y}) \quad (\spadesuit)$$

where $\bar{q}$ is the quaternion conjugate to q .

We consider the vector bundle over $\mathbb{H}^2$ whose fiber over $(\mathbf{x}, \mathbf{y})$ is $E_{\mathbf{x}, \mathbf{y}} \equiv \ker \Delta(\mathbf{x}, \mathbf{y})$. By Eq. $(\spadesuit)$ it descends to a rank- N complex vector bundle $E \rightarrow \mathbb{P}^1(\mathbb{H})$ over the *quaternionic projective line* $\mathbb{P}^1(\mathbb{H})$ which is diffeomorphic to the 4-sphere S^4 . The punctured 4-sphere is

$$\mathbb{R}^4 \simeq S^4 \setminus (\text{South pole}) \equiv \mathbb{P}^1(\mathbb{H}) \setminus (\mathbf{1} := \mathbf{0})$$

$E|_{\mathbb{R}^4}$ is the vector bundle associated to the instanton connection A via the fundamental representation. By construction both E and A extends smoothly to the compactification $S^4 \equiv \mathbb{P}^1(\mathbb{H})$ of $\mathbb{R}^4$. This shows that the instanton has *finite* action. The matrix $\Delta(\mathbf{x}, \mathbf{y})$ acts on $\mathbb{C}^{N+2k}$ which can be seen as the fiber of a trivial bundle over $\mathbb{P}^1(\mathbb{H})$. Hence we have the exact bundle sequence

$$0 \rightarrow E \rightarrow \mathbb{C}^{N+2k} \rightarrow \text{Im } \Delta(\mathbf{x}, \mathbf{y}) \rightarrow 0.$$

So the Pontryagin number of E is

$$p(E) = -p(\text{Im } \Delta(\mathbf{x}, \mathbf{y})) = -k.$$

Since $\text{Im } \Delta(\mathbf{x}, \mathbf{y})$ is isomorphic to k copies of the tautological bundle $\mathcal{T} \rightarrow \mathbb{H}\mathbb{P}^1$ with Pontryagin number 1, $\mathcal{T}$ is the Swann bundle over the quaternionic Kähler symmetric space $\mathbb{H}\mathbb{P}^1$ [23]

- (i) takes values in $\mathfrak{su}(N)$;
- (ii) is anti-self-dual;
- (iii) has finite Yang–Mills action $\int \text{tr}(F_{\mu\nu} F^{\mu\nu}) < \infty$;
- (vi) has topological charge (Pontryagin number) k .

All instanton solutions of $SU(N)$ YM have the form (14.49) for some ADHM data.

Proof (i) The second equality in (14.49) shows that A takes values in $\mathfrak{u}(N)$. The trace part vanishes by (ii), (iii), and the fact that there is no Abelian anti-self-dual field configuration in $\mathbb{R}^4$ with finite action. (ii) From Eq. (14.48) $\Delta \partial_\mu U = -(\partial_\mu \Delta)U$, and then the field-strength becomes

$$\begin{aligned}
F_{\mu\nu} &= \partial_\mu(U^\dagger \partial_\nu U) + (U^\dagger \partial_\mu U)(U^\dagger \partial_\nu U) - (\mu \leftrightarrow \nu) = \\
&= (\partial_\mu U^\dagger)(\partial_\nu U) - (\partial_\mu U^\dagger)U U^\dagger(\partial_\nu U) - (\mu \leftrightarrow \nu) = \\
&= (\partial_\mu U^\dagger)P(\partial_\nu U) - (\mu \leftrightarrow \nu) = (\partial_\mu U^\dagger)\Delta^\dagger(\mathbf{1}_2 \otimes f)\Delta(\partial_\nu U) - (\mu \leftrightarrow \nu) = \\
&= U^\dagger[(\partial_\mu \Delta^\dagger)(\mathbf{1}_2 \otimes f)(\partial_\nu \Delta) - (\mu \leftrightarrow \nu)]U = 2U^\dagger \begin{pmatrix} 0 & 0 \\ 0 & \sigma_{\mu\nu} \otimes f \end{pmatrix} U
\end{aligned} \tag{14.50}$$

where $\sigma_{\mu\nu} \equiv \frac{1}{2}(\sigma_\mu \bar{\sigma}_\nu - \sigma_\nu \bar{\sigma}_\mu)$ is the generator of $SO(4)$ rotations acting on Euclidean Weyl fermions of chirality $+1$, which is anti-self-dual (Box 14.3). This shows that

$$F_{\mu\nu} = -\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{\rho\sigma}. \tag{14.51}$$

(iii) The *classical*⁸ 4d Yang–Mills theory is conformal invariant. Since $\mathbb{R}^4$ is conformal equivalent to the punctured 4-sphere $S^4 \setminus (\text{North pole})$, it is sufficient to show that the gauge configuration A seen as a connection on $S^4 \setminus (\text{North pole})$ extends smoothly over the North pole, yielding a smooth gauge connection on the compact space $S^4 \simeq \mathbb{P}^1(\mathbb{H})$, the projective line over the quaternions. The extension over S^4 is constructed in Box 14.4. (iv) Let $E \rightarrow S^4$ the vector bundle associated to A via the fundamental representation of $SU(N)$. In Box 14.4 it is shown that its Pontryagin number is $-k$. Finally, the statement that all instantons are of this form is equivalent to saying that the moduli space of instantons coincides with the moduli space of ADHM data, a point already discussed (from a physical perspective) in Sect. 12.8. $\square$

It is convenient to rephrase Eq. (14.49) in more intrinsic terms. As better explained in Box 14.4, the instanton field A is a connection on a rank- N subbundle $E \rightarrow \mathbb{P}^1(\mathbb{H})$ of the trivial bundle of fiber $\mathbb{C}^{N+2k}$ with inclusion map

$$U := E \rightarrow \mathbb{C}^{N+2k} \times \mathbb{P}^1(\mathbb{H}). \tag{14.52}$$

The trivial bundle carries the natural trivial connection d . Let $\nabla^E := E \rightarrow E \otimes \Omega^1$ be the connection on the subbundle E induced by the trivial connection: by definition [24]

$$U \nabla^E \psi = (1 - P)d(U\psi) = UU^\dagger d(U\psi) = U(d\psi + (U^\dagger dU)\psi) \tag{14.53}$$

so that *the instanton gauge field A is the subbundle connection ∇^E on $E \equiv \text{im } U$ induced by the trivial connection on $\mathbb{C}^{N+2r} \times \mathbb{P}^1(\mathbb{H})$.*

We wish to understand this result in terms of the physics of BPS branes.

The ADHM Construction in Terms of D-Branes

We know that a system of k D(−1) and N (Euclidean) D3 realizes a $SU(N)$ k -instanton along the Higgs branch $\mathcal{M}_{N,k}$. More generally, we can T -dualize some orthogonal directions and consider a system of k D p and N D($p + 4$) parallel branes. On the D($p + 4$) we have a $(p + 5)$ -dimensional gauge field configuration which looks as an instanton in the four directions orthogonal to the D p 's and is translational invariant in the $(p + 1)$ parallel ones.

We need a “physical device” which measures the gauge field A_μ living on the stack of $(p + 4)$ -branes along the system's Higgs branch. This device can only be a

⁸ At the quantum level the conformal symmetry is anomalous: the 4d YM β -function is non-zero.

Table 14.1 The light degrees of freedom propagating on the world-sheet of a probe D1-brane in presence of k D5's and N D9's. **Caution:** this brane configuration is unphysical: its actual meaning and purpose is explained in the main text

Sector	Bosons	Right-moving fermions	Left-moving fermions
1–1	$a_\mu, \quad X^{\alpha\dot{\alpha}}, \quad X^{ai}$	$\tilde{\psi}^{\alpha\dot{\alpha}}, \quad \tilde{\psi}^{ai}$	$\psi^{\alpha\dot{\alpha}}, \quad \psi^{ai}$
1–5 & 5–1	$\phi^{a\bar{m}}, \quad \bar{\phi}^{am}$	$\tilde{\chi}^{\alpha\bar{m}}, \quad \tilde{\chi}^{\alpha m}$	$\chi^{\dot{\alpha}\bar{m}}, \quad \bar{\chi}^{\dot{\alpha} m}$
1–9 & 9–1			$\lambda^M, \quad \bar{\lambda}^M$

probe brane. By a “probe” we mean an auxiliary physical system whose dynamics is very sensitive to the configuration of the relevant system we wish to observe, but has a negligible back-reaction on this configuration. This means that the fields living on the Dp and $D(p+4)$ branes should enter the world-volume QFT of the probe brane as couplings.

Consider the *formal* system of k D5's and N D9's in Type IIB.⁹ This system is unphysical since nothing cancels the 10-form tadpole (cf. Sect. 5.3). However we use it just a technical trick to simplify the kinematics: the system we really have in mind is the configuration obtained by T -dualizing one (or two) normal direction(s), getting a brane configuration which makes perfect physical sense. The formulae for the physical system are most easily obtained by dimensional reduction of the formal one [26].

To the (formal) system of k D5's and N D9's we add a D1 parallel to the D5's. From the D1 viewpoint, the d.o.f. living on the D5's and D9's are infinitely heavy, hence decoupled (cf. Sect. 12.6.2), and there is no back reaction on them. Thus the D1 is a genuine probe. It projects to a point in the transverse $\mathbb{R}^4$ where the instanton lives, so it can scan the instanton field at each point of the relevant $\mathbb{R}^4$. This brane set-up breaks the 10d Lorentz group to the subgroup¹⁰

$$SO(9, 1) \rightsquigarrow SO(1, 1) \times SO(4)_\perp \times SO(4)_\parallel \sim SO(1, 1) \times \left(\prod_{A=1}^4 SU(2)_A \right). \quad (14.54)$$

We write $\alpha = 1, 2, \dot{\alpha} = 1, 2, a = 1, 2$ and $i = 1, 2$ for the indices of the four $SU(2)$'s.

The light d.o.f. living on the D1 are the ones arising from the 1–1 strings (see Sect. 13.1) plus the ones from the light 1–9, 1–5, 9–1, 5–1 strings. Just as in Sect. 13.4 the 1–9 plus 9–1 produce complex left-moving fermions λ^M ($M = 1, \dots, N$) but, in contrast to the Type I case, now we do *not* impose the Ω -projection. The 1–5 plus 5–1 strings produce a full hypermultiplet in the fundamental of $U(k)$; cf. Sect. 12.6.2. The resulting light dynamical d.o.f. propagating on the D1 world-sheet are listed in Table 14.1.

⁹ In this paragraph we follow [25, 26].

¹⁰ Here the symbols $\parallel$ and $\perp$ refer to directions respectively parallel and orthogonal to the D5's.

The fields living on the D5's are numerical coupling constants from the D1 perspective. The 5–9 and 9–5 scalars $h_M^{\alpha m}$ form a full hypermultiplet in the bi-fundamental of $U(k) \times U(N)$. The 5–5 scalars give the D5s' transverse position matrix $Y^{\alpha\dot{\alpha}}$ which we see as a $k \times k$ matrix with quaternionic entries. In Sect. 12.8 the v.e.v. of these scalars were identified with the ADHM data $(B_1, B_2, Q, \tilde{Q})$

$$Y^{\alpha\dot{\alpha}} \equiv \begin{pmatrix} B_2 & B_1 \\ -B_1^\dagger & B_2^\dagger \end{pmatrix}, \quad h^{1m}_M \equiv Q^m_M, \quad h^{2m}_M \equiv \tilde{Q}^m_M. \quad (14.55)$$

In absence of other branes the D1 world-sheet theory is invariant under a (8, 8) SUSY. In presence of D5's the D1 supersymmetry reduces to (4, 4) (cf. Sect. 12.6), while in presence of D9's to (0, 8) (cf. Sect. 13.4). The full D1–D5–D9 configuration is invariant under the intersection of the SUSYs preserved by the various subsectors, i.e. by a (0, 4) two-dimensional supersymmetry.

We consider the 2d effective action on the world-sheet of a D1 probing our D5/D9 brane system. The couplings involving only fields from one or two p - p' sectors can be read from the analysis of the D1–D5 and D1–D9 systems (see Sects. 12.6.2 and 13.4). There remains to fix the coupling involving fields from the three sectors 1–5, 5–9, and 9–1. These three-sector couplings are string amplitudes on the upper half-plane with three massless vertex insertion on the boundary at 0, 1, and ∞ , with boundary conditions given by the D1 on the boundary segment $(-\infty, 0)$, the D5 on $(0, 1)$ and the D9 on $(1, \infty)$. The inserted vertex operators correspond to the vacua of the corresponding p - p' string. The amplitude is clearly non-zero, producing (in particular) a Yukawa coupling¹¹

$$\lambda^M \tilde{\chi}_m^\alpha h_{\alpha M}^m. \quad (14.56)$$

All other couplings are obtained from these ones by (0, 4) supersymmetry. By translational invariance the fields $Y^{\alpha\dot{\alpha}}$ and $X^{\alpha\dot{\alpha}}$ may enter in the effective Lagrangian only in the combination $Y^{\alpha\dot{\alpha}} - X^{\alpha\dot{\alpha}}$. We stress that in order to have (0, 4) SUSY on the D1 world-sheet the four matrices B_1, B_2, Q , and $\tilde{Q}$ should satisfy the equations (14.42), i.e. should be a valid ADHM datum. Indeed, we saw in Sect. 12.8 that these equations are precisely the condition for the D5–D9 system we are probing to be a quarter-BPS state which preserves 8 SUSYs. In other words: *the relevant couplings on the D1 probe world-sheet are specified by an ADHM datum.*

We combine the left-moving fermions into a single vector with $N + 2k$ components

$$\lambda^A = (\lambda^{\tilde{M}}, \chi^{\dot{1}\tilde{m}}, \chi^{\dot{2}\tilde{m}}), \quad A = 1, 2, \dots, N + 2k. \quad (14.57)$$

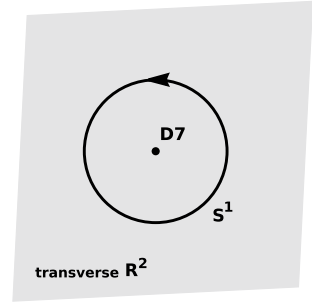
The terms in the effective Lagrangian which contain the fermions λ^A are¹²

$$L = \bar{\lambda}^A \bar{\partial} \lambda^A + \tilde{\chi}^{\alpha m} \Delta_{\alpha m, A} \lambda^A + \text{h.c.} + \dots$$

¹¹ We absorb the overall coefficient in the normalization of the fields.

¹² We omit the coupling of λ^A with the 2d Abelian gauge field a_μ since it is non-dynamical, see the corresponding statement in Sect. 13.1.

Fig. 14.2 The spatial plane $\mathbb{R}^2$ transverse to a D7-brane (located at the origin) with a circle S^1 enclosing it. The $F^{(1)}$ magnetic flux through S^1 measures the RR charge of the D7-brane.



where the $2k \times (N + 2k)$ matrix of 2d Yukawa couplings $\Delta_{am,A}$ is the matrix Δ in Eq. (14.44). Now we are almost done. Let us see what the above Lagrangian L becomes at low-energy along the Higgs branch where the Yukawa coupling matrix Δ has maximal rank $2k$. $2k$ out of the $N + 2k$ fermions λ^A get mass and can be integrated away. The linear combinations which remain massless form the subbundle E of the trivial bundle $\mathbb{C}^{2k+N}$ over $\mathbb{H} \simeq \mathbb{R}^4$. Using the inclusion map $U := E \rightarrow \mathbb{C}^{2k+N} \times \mathbb{H}$, the light fermions can be written as $\lambda^A = U^A_m \zeta^m$ for ζ^m ($m = 1, \dots, N$) unconstrained Fermi fields. The low-energy effective kinetic terms then become

$$\bar{\zeta} U^\dagger \bar{\partial}(U \zeta) = \bar{\zeta} (\bar{\partial} + U^\dagger \partial_\mu U \bar{\partial} X^\mu) \zeta. \quad (14.58)$$

The world-sheet connection is the pull-back of the gauge field A_μ in the $\mathbb{R}^4$ factor orthogonal to the D5's which is precisely the instanton profile we are probing with the D1. Then $A_\mu = U^\dagger \partial_\mu U$ which is the ADHM result (14.49). Note that, by construction,¹³ the gauge connection is the sub-bundle connection on the subbundle E of fermions which are massless along the Higgs branch. From this point of view the fact that the field strength of A_μ is *anti-self-dual* is a direct consequence of 2d supersymmetry together with the $Spin(4)$ chirality of the left-moving fermions; cf. Eq. (14.57).

The brane construction gives a physical understanding of the ADHM construction originally obtained from Algebro-Geometric and twistor arguments [18–20].

14.3 The Idea of F-Theory

In this section, we present a general picture of the non-perturbative completion of Type IIB called F-theory. F-theory was introduced by Cumrun Vafa in [27] (see also [28, 29]). For a review see, e.g. [30].

¹³ That is: as a standard consequence of the procedure of integrating out massive stuff.

The first observation is that the massless complex scalar of 10d Type IIB

$$\tau = C^{(0)} + ie^{-\Phi} \quad (14.59)$$

need *not* to be continuous *nor* univalued in a physically regular state. Indeed, consider a D7-brane whose transverse directions span a 2-plane (see Fig. 14.2). The D7-brane is a magnetic source for the gauge scalar $C^{(0)}$ with unit magnetic charge given by the flux of its field-strength $F^{(1)} \equiv dC^{(0)}$ on the S^1 at spatial infinity

$$1 \equiv Q_m = \int_{S^1} F^{(1)} \equiv \int_{S^1} dC^{(0)}. \quad (14.60)$$

This equation shows that the RR field-strength $F^{(1)}$ is a closed 1-form which is *not* exact: rather $F^{(1)}$ represents an *integral cohomology* class:

$$[F^{(1)}] \in H^1(M, \mathbb{Z}). \quad (14.61)$$

The last statement is just Dirac quantization of the magnetic charge $Q_m \in \mathbb{Z}$ whose dual electric charge is the integral D(−1) instanton number. This observation implies that the RR potential $C^{(0)}$ is a *periodic scalar* (i.e. it is $(2\pi)^{-1}$ times an angle¹⁴),

$$C^{(0)} \sim C^{(0)} + 1, \quad (14.62)$$

so, in a physically regular configuration, $C^{(0)}$ must be continuous/univalued only mod 1. We may equivalently say that $T := \tau \rightarrow \tau + 1$ is a *discrete gauge symmetry*, meaning a redundancy in the description which leaves invariant all physical quantities.

However the discrete gauge symmetry which acts on τ cannot be just the Abelian group ($\simeq \mathbb{Z}$) generated by T , and so the discontinuous jumps of the scalar field τ cannot consist only in integral shifts. We repeat the argument which we presented before in a wider context (see the paragraph after Note 13.2). We recall the

Fact 14.1 In any physical theory the gauge group $\mathcal{G}$ is a *normal* subgroup of the group of all symmetries $\mathbf{G}$

$$\mathcal{G} \triangleleft \mathbf{G}. \quad (14.63)$$

Proof By Gauss' law, we have $h|\psi\rangle = |\psi\rangle$ for all elements $h \in \mathcal{G}$ and all physical states $|\psi\rangle$. Since $\mathbf{G}$ is a symmetry, $g|\psi\rangle$ is also a physical state for all $g \in \mathbf{G}$. Then

$$g^{-1}hg|\psi\rangle = g^{-1}g|\psi\rangle = |\psi\rangle, \quad (14.64)$$

so $g^{-1}hg$ acts as the identity on all physical states, i.e. $g^{-1}\mathcal{G}g \subseteq \mathcal{G}$ for all $g \in \mathbf{G}$. $\square$

In Lemma 5.1 we saw that the normal closure of $T^{\mathbb{Z}} \subset SL(2, \mathbb{Z})$ is the full $SL(2, \mathbb{Z})$. Since $SL(2, \mathbb{Z})$ is a symmetry of Type IIB (cf. Sect. 13.1) we conclude

¹⁴ In Sect. 13.1.2 we saw that the θ -angle of the SYM living on the D3 world-volume is $\theta \equiv 2\pi \langle C^{(0)} \rangle$.

that the full duality group $SL(2, \mathbb{Z})$ is a *discrete gauge symmetry* of Type IIB and in particular the scalar τ needs to be univalued only modulo $SL(2, \mathbb{Z})$, that is, in Type IIB the massless scalar τ takes value in the space

$$\mathcal{H}/SL(2, \mathbb{Z}) \equiv \text{moduli space of elliptic curves} \simeq \mathbb{P}^1 \setminus \{\infty\}, \quad (14.65)$$

where the isomorphism $\overline{\mathcal{H}/SL(2, \mathbb{Z})} \rightarrow \mathbb{P}^1$ is given by $\tau \mapsto j(\tau)$ with $j(\tau)$ the usual modular invariant function (see Box 9.4). Two elliptic curves are isomorphic if and only if their periods have the same *modular invariant* $j(\tau)$.

In conclusion: in Type IIB we cannot associate to a point p of the 10d spacetime manifold M_{10} a well-defined value $\tau(p)$ of the complex scalar τ , but we can associate to $p \in M_{10}$ the modular invariant $j(\tau(p))$ or, equivalently, an elliptic curve E_p well-defined up to isomorphism. In absence of 3- and 5-form fluxes, a bosonic configuration is most naturally seen as a fibration

$$\pi := X_{12} \rightarrow M_{10} \quad (14.66)$$

of a 12-dimensional space X_{12} over the “physical spacetime” M_{10} whose fiber over a *generic* point $p \in M_{10}$ is the elliptic curve of modular invariant $j(\tau(p))$. Non-generic singular fibers appear over submanifolds $D_i \subset M_{10}$ of real codimension $\ell \geq 2$ which are the support of $(9 - \ell)$ -branes of various kinds. The locus $D \subset M_{10}$ with singular fibers is called the *discriminant*.

Before going to F-theory we make a couple of remarks about elliptic fibrations.

Remark 14.1 There is a subtle difference between a fibration in elliptic curves and a fibration in genus-1 curves. An elliptic curve is a complex curve with an Abelian group structure. It is given by a genus-1 curve C together with the choice of one point $0 \in C$ which plays the role of the neutral element for the group law [31]. A fibration π in elliptic curves has (at least) one global section,¹⁵ namely the zero section which associates to a point $p \in M_{10}$ the zero element of its group fiber E_p . On the contrary a general fibration in genus-1 curves may have no section. For all fibration in genus-1 curves, there is a fibration in elliptic curves with the same function $p \mapsto j(\tau(p))$ (the Kodaira functional invariant [37, 38]): in Kodaira’s language this is the *Jacobian fibration* [37–39]. If we are only interested in the configuration of the metric and τ we may (and do!) assume π to have a global zero section. See however [40] for F-theory in the more general set-up of fibrations in genus-1 curves.

Remark 14.2 The elliptic fibration (with zero section) $\pi := X_{12} \rightarrow M_{10}$ contains more information than just the 10d space-time M_{10} and the function $j(\tau) := M_{10} \rightarrow \mathbb{P}^1$ which specifies the scalar field configuration. Indeed there are inequivalent fibrations with the same functional invariant $j(\tau)$ (cf. discussion in Box 14.5).

¹⁵ The group of all global section is called the *Mordell–Weil group* [32–34]. The structure of the Mordell–Weil group is an important physical datum of an F-theory configuration [35, 36].

F-Theory

In F-theory [27] one sees the 12-dimensional space X_{12} as the fundamental physical manifold. The IIB supergravity degrees of freedom propagate only along the zero-section of π which is a copy of M_{10} embedded in X_{12} . This is analog to gauge degrees of freedom propagating only along the world-volume of a D-brane inside the stringy 10d space-time M_{10} . In F-theory the zero-section $M_{10} \subset X_{12}$ is sometimes viewed as the *gravitational brane*. In addition we have lower-dimensional gauge branes where the YM d.o.f. live.

Since gravity propagates only along the gravitational brane $M_{10} \subset X_{12}$, the metric in the fibers has no physical meaning. The fibers have an intrinsic complex structure given by the modular invariant $j(\tau)$ but not an intrinsic metric.¹⁶

There are several good reasons to believe that the 12-dimensional viewpoint is the deep and intrinsic one. Let us consider a compactification to $2n < 10$ dimensions where the scalar field τ is non-constant (for simplicity we assume the 3- and 5-form fluxes to vanish). We ask when such a configuration preserves some supersymmetries. In terms of the 12-dimensional geometry X_{12} the answer is simple and elegant:

Claim 14.3 *In absence of 3-/5-form fluxes, in order to preserve some supersymmetry, X_{12} must have the form*

$$\mathbb{R}^{2n-1,1} \times K_{6-n} \quad (14.67)$$

with K_{6-n} an elliptic Calabi–Yau $(6-n)$ -fold with section. If the CY is strict, the low-energy effective theory is a $\mathcal{N} = 1$ SUGRA in $d = 2n$.

A Calabi–Yau manifold K is called *elliptic* (with section) if it admits a holomorphic fibration (with section) $K \rightarrow B$ over some complex base B whose generic (smooth) fibers are elliptic curves. The gravitational brane then is $M_{10} = \mathbb{R}^{2n-1,1} \times B$, and the profile of the scalar field τ on M_{10} is given by the period of the elliptic fibers.

If $6 - n > 2$ and K is a strict elliptic CY, the base B is automatically an algebraic variety (in fact a Fano one). For a strict elliptic CY the meromorphic function $j := B \rightarrow \mathbb{P}^1$ is non-constant,¹⁷ hence surjective, and there are regions of the 10d “space-time” M_{10} where the string coupling $1/\text{Im } \tau$ is strong: i.e. all BPS configuration with non-constant τ is *inherently strongly coupled* and cannot be described in string perturbation theory. F-theory is the non-perturbative completion of IIB which allows us to study these situations in *exact terms* for the same reasons as in the Calabi–Yau compactifications described in Chap. 11. There are several interesting phenomena that cannot happen in string perturbation theory yet present in the wider F-theory context: exceptional gauge groups, “exotic” representations, etc. Some of these non-perturbative mechanisms are relevant for model building of GUT-like low-energy

¹⁶ For some computations it is technically convenient to endow the fibers with some conventional metric. This metric has no physical content and it is used as a mere regulator. One writes it in the form ϵh_{ij} and takes the limit $\epsilon \rightarrow 0$ at the end of the computation.

¹⁷ If $j(\tau) \equiv c$ is constant, there is a finite cover K' of K of the form $B' \times E$ for a fixed elliptic curve E with $j(E) = c$ and B' a Calabi–Yau. Hence the holonomy Lie algebra is contained in $\mathfrak{su}(5-n)$ and K is not strict.

BOX 14.5 Homological invariant and Kodaira monodromies

The present discussion is similar to the one in Box 11.1 except that here we have a *smooth* family of elliptic curves $X \rightarrow \dot{M} \equiv M \setminus D$ instead of a *holomorphic* one. A smooth path $\gamma := [0, 1] \rightarrow \dot{M}$ again defines a diffeomorphism between the fibers $\phi_\gamma := X_{\gamma(0)} \rightarrow X_{\gamma(1)}$ whose homotopy class depends only on the homotopy class of γ . This gives a map

$$\mu := \pi_1(M \setminus D, *) \rightarrow \text{MCG}(X_*) \simeq SL(2, \mathbb{Z}), \quad \mu := [\gamma] \mapsto [\phi_\gamma(1)]$$

which coincides with the monodromy representation on $H_1(X_*, \mathbb{Z})$. The modular action of $\pi_1(M \setminus D, *)$ on τ factors through the quotient $PSL(2, \mathbb{Z}) \equiv SL(2, \mathbb{Z})/\mathbb{Z}_2$; the functional invariant $j := \dot{M} \rightarrow \mathbb{C}$ is sensitive only to the projective representation in $PSL(2, \mathbb{Z})$. Families over $\dot{M}$ with inequivalent monodromies μ with the same image in $PSL(2, \mathbb{Z})$ have the same functional invariant but are not isomorphic.

In the interesting cases the family is holomorphic and D is a divisor with irreducible components D_i . If M is simply-connected, $\pi_1(M \setminus D)$ is generated by lassos γ_i encircling each D_i . The intrinsic properties of the 7-brane whose support is D_i are determined by the conjugacy class of the local monodromy $\mu(\gamma_i)$ at D_i in $SL(2, \mathbb{Z})$.

Kodaira exceptional fibers The local monodromies which may appear in codimension-1 in a holomorphic elliptic fibration and the dependence of the functional invariant j on a local coordinate z such that D_i has local equation $z = 0$ were classified by Kodaira [37, 38, 46]. The allowed local monodromies are the elements $\rho \in SL(2, \mathbb{Z})$ which are quasi-unipotent and satisfy the non-negativity condition of the nilpotent orbit theorem [47]

$$(\rho^m - 1)^2 = 0 \quad \text{and} \quad -i \operatorname{tr}(\sigma_2(\rho^m - 1)) \geq 0, \quad \text{for some } m \in \mathbb{N}.$$

The fibers types and the corresponding monodromy conjugacy classes are listed in Table 14.2. I_0 is a regular elliptic fiber.

theories which lead to realistic models of the real world “beyond the Standard Model” [41–44] (for a review see [45]).

Claim 14.3 can be easily verified directly, by writing down the first-order differential equations for a BPS configuration in Type IIB SUGRA with τ non-constant. We prefer to prove it by duality with M-theory compactified on the same Calabi–Yau manifold, a result of independent interest (see Sect. 14.3.1 below).

Note 14.5 The fact that we have a SUSY theory in 12 dimensions may seem in contradiction with our statement in Chap. 8 that the highest dimension in which we may have a SUSY model is 11. The point is that in Chap. 8 we *assumed* spacetime to be a pseudo-Riemannian manifold with a metric of Lorentzian signature $(d - 1, 1)$, i.e. with a *single* time direction. The statement that we have at most 32 supercharges is correct in any signature, but the number of real components of an irreducible spinor is signature (r, s) (with $r \geq s$) is (cf. Table 2.1 in [48])

$$2^s \mathbf{N}(r - s) \equiv 2^{s-1} N(r - s + 2) \quad (14.68)$$

where $\mathbf{N}(k)$ is the function defined in Eq. (2.57) of [48], equivalently $N(k)$ is the function defined in our Eq. (8.23). Thus we may have 32 supercharges in 12d provided the signature is (10, 2) or (6, 6). The signature of the metric in X_{12} is not defined, since there is no metric except along the sub-manifold M_{10} . If one insists to endow X_{12} with some metric, one easily sees that the most natural signature is (10, 2) [27] which is consistent with a 32-supercharge supersymmetry.

4-Form Flux

The two 3-form field-strengths of Type IIB, $H^{(3)}$ and $F^{(3)}$, transform as a doublet under $SL(2, \mathbb{Z})$; cf. Sect. 8.5. This means that when we go along a non-trivial loop¹⁸ $\gamma \subset M_{10} \setminus \cup_i D_i$ we get back with fluxes $H^{(3)}$, $F^{(3)}$ rotated by the monodromy $\rho(\gamma) \in SL(2, \mathbb{Z})$ see Boxes 11.1 and 14.5. Hence $H^{(3)}$ and $F^{(3)}$ are not well-defined (unvalued) 3-forms in a F-theory background. We need to replace them by a globally well-defined flux-form living on the proper 12d manifold X_{12} . The natural $SL(2, \mathbb{Z})$ invariant is the 4-form on the total space X_{12}

$$G^{(4)} = dy \wedge H^{(3)} - dx \wedge F^{(3)}, \quad (14.69)$$

where the differentials $\{dx, dy\}$ give a symplectic basis of $H^1(E_p, \mathbb{Z})$ on the elliptic fiber E_p .

Seven-Branes

The locus of “bad” fibers $S \subset B$ is a divisor with irreducible components S_i . The submanifolds

$$D_i \equiv \mathbb{R}^{2n-1,1} \times S_i \subset M_{10} \equiv \mathbb{R}^{2n-1,1} \times B \quad (14.70)$$

are the support of 7-branes on which the gauge d.o.f. lives. There are several kinds of 7-branes depending on the way the fiber degenerates along S_i . Besides the perturbative IIB possibilities, stacks of D7’s and O7’s, there are some new ones which exist only at the non-perturbative F-theory level. The gauge group of the theory living on each 7-brane depends on its singular fiber type: see next paragraph for details. Over the double intersections $S_i \cap S_j$ the fiber gets even more singular producing additional “matter” d.o.f. which are charged with respect to the gauge d.o.f. living on the branes D_i and D_j and higher codimensional singularities correspond to couplings (see [30, 42]).

Exceptional Fibers & Gauge Groups The singular fibers which may appear in codimension-1, that is, over the generic point of the i -th irreducible component S_i , were classified by Kodaira [37, 38, 46] (see also [39]). Let ℓ_i be a small circle which goes around the 7-brane on $D_i = \mathbb{R}^{2n-1,1} \times S_i$: see Fig. 14.2 for the D7 example. Going around the circle the complex scalar τ gets back to itself up to a modular transformation by an element $\gamma_i \in SL(2, \mathbb{Z})$:

¹⁸ Here the D_i ’s are the irreducible components of the discriminant D of pure complex codimension 1 (in a geometry of the form (14.67)). Then $M_{10} \setminus \cup_i D_i$ is the *open* domain of “good” points in the spacetime M_{10} , i.e. points whose fiber is a smooth elliptic curve.

Table 14.2 Kodaira's exceptional elliptic fibers in codimension 1 (here $n = 0, 1, 2, \dots$)

Kodaira fiber	$\mathbf{I}_n$	$\mathbf{I}_n^*$	$\mathbf{II}$	$\mathbf{II}^*$	$\mathbf{III}$	$\mathbf{III}^*$	$\mathbf{IV}$	$\mathbf{IV}^*$
Lie algebra $\mathfrak{g}$	$\mathfrak{u}(n)$	$\mathfrak{so}(2n+8)$	—	E_8	$\mathfrak{su}(2)$	E_7	$\mathfrak{su}(3)$	E_8
$j(\tau)/(12)^3$	order n pole	order n pole	0	0	1	1	0	0
Monodromy	$\begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -1 & -n \\ 0 & -1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ -1 & -1 \end{bmatrix}$	$\begin{bmatrix} -1 & -1 \\ 1 & 0 \end{bmatrix}$
Euler no.	n	$n+6$	2	10	3	9	4	8

$$\tau' = (a\tau + b)/(c\tau + d), \quad \gamma_i = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (14.71)$$

The singular fiber over the generic point of D_i is determined by the conjugacy class $[\gamma_i]$ in $SL(2, \mathbb{Z})$ see Box 14.5 and Table 14.2. For instance when $\gamma_i = T^n$ (type $\mathbf{I}_n$)

$$\tau' = \tau + n \iff \oint_{S^1} dC^{(0)} = n, \quad (14.72)$$

the 7-brane carries n units of RR charge and may be identified with a stack of n D7 branes. When $\gamma_i = -1$ (type $\mathbf{I}_0^*$) the 3-form field strengths $F^{(3)}$, $H^{(3)}$ get reflected, and the 7-brane should be an orientifold. It cannot be simply a O7 plane since τ is unvalued near a type $\mathbf{I}_0^*$ fiber, which then carries *zero* RR charge, whereas O7 has RR charge -4 (cf. Eq. (12.26)); therefore a $\mathbf{I}_0^*$ 7-brane is naturally interpreted as a O7 with four D7 on top of it. Likewise, a 7-brane of type $\mathbf{I}_n^*$ (monodromy $-T^n$) carries the conserved charges of $(n+4)$ D7 on top of a O7. The fibers of Kodaira types

$$\mathbf{II}, \mathbf{II}^*, \mathbf{III}, \mathbf{III}^*, \mathbf{IV}, \mathbf{IV}^* \quad (14.73)$$

represent new non-perturbative 7-branes with no counterpart in perturbative IIB theory. This is already clear from Table 14.2: at string weak coupling $\tau \rightarrow i\infty$ $j(\tau) \rightarrow \infty$ and this is consistent only with the ‘perturbative’ Kodaira types $\mathbf{I}_n, \mathbf{I}_n^*$.

Gauge Groups On each 7-brane live the d.o.f. of a supersymmetric gauge theory just as in perturbative IIB. The gauge group depends on the Kodaira type of the corresponding generic fiber. As in the case of the gauge groups living on 6-branes in M-theory (Sect. 14.1.1), the assignment of a gauge Lie algebra to a fiber singularity type goes through the McKay correspondence. As discussed in Box 14.6, a singular Kodaira fiber is a bouquet of spheres with intersection form the negative of an affine Cartan matrix of type $\mathfrak{g} \in ADE$. The gauge Lie algebra of the SYM on the 7-brane is the same $\mathfrak{g}$: see Table 14.2. For the ‘perturbative’ types $\mathbf{I}_n$ and $\mathbf{I}_n^*$ we recover the gauge groups $U(n)$ and $SO(2n+8)$ we got in Chap. 12 from a weak coupling analysis. The exceptional gauge Lie groups E_6 , E_7 , and E_8 are realized only in the F-theory and have no perturbative analogue.

Note 14.6 F-theory has also other 7-branes associated to “frozen” singular fibers [16, 49]. They are the F-theory counterpart of the subtler orientifold planes in perturbative IIB mentioned in Note 12.2 which are related to toroidal compactifications without vector structure [50], as well as of the “frozen” singularities of M-theory in Note 14.4.

Weierstrass Model

A holomorphic elliptic fibration $X \rightarrow B$ over the complex base B , with section, has a¹⁹ birational *Weierstrass model*. A Weierstrass model of an elliptic fibration (with section) over the complex base B is specified by the data $(\mathcal{L}, a, b)$ where $\mathcal{L} \rightarrow B$ is a holomorphic line bundle and a (resp. b) is a holomorphic section of $\mathcal{L}^4$ (resp. $\mathcal{L}^6$). The Weierstrass model of the elliptic fibration X is the hypersurface in the total space of the bundle $\mathcal{L}^2 \oplus \mathcal{L}^3$ of equation

$$y^2 = x^3 + ax + b \quad (14.74)$$

where (x, y) are fiber coordinates in $\mathcal{L}^2 \oplus \mathcal{L}^3$. The equation is well defined since both sides are sections of $\mathcal{L}^6$. The *discriminant* is the divisor of the discriminant of the cubic polynomial in the RHS of (14.76)

$$D \equiv \{\Delta \equiv 4a^3 + 27b^2 = 0\} \subset B. \quad (14.75)$$

The fiber X_p over a point p in the basis B , which does not belong to the discriminant D , is the smooth elliptic curve of equation

$$y^2 = x^3 + a(p)x + b(p), \quad (x, y) \in \mathbb{C}^2, \quad \Delta(p) \neq 0 \quad (14.76)$$

where $a(p), b(p)$ are the evaluation at p of the sections a, b in a trivialization of $\mathcal{L}$. The curve X_p is independent of the chosen trivialization up to isomorphism. Indeed the functional invariant $j := B \rightarrow \mathbb{P}^1$ is the well-defined meromorphic function

$$j = (24)^3 \frac{4a^3}{4a^3 + 27b^2} \quad (14.77)$$

which may have a pole only at zeros of Δ . Δ is a section of $\mathcal{L}^{12}$ i.e.

$$O(\Delta) = \mathcal{L}^{12}. \quad (14.78)$$

In the case of an F-theory compactification preserving some symmetry the elliptic fibration X is also a Calabi–Yau space, that is, the canonical bundle K_X is trivial. The holomorphic differential along the fiber is dx/y , which is a section of $\mathcal{L}^{-1}$. A top degree holomorphic differential has the local form

¹⁹ The Weierstrass model is obtained by blowing down all components of the fibers which do not cross the zero section.

BOX 14.6 Kodaira fibers

We consider a smooth complex surface S elliptically fiber over a complex curve B . An exceptional (non-smooth) fiber S_b over $b \in D \subset B$ is an effective divisor of the form

$$S_b \equiv \pi^{-1}(b) = \sum_i m_i \Theta_i$$

where $\Theta_i \subset S$ are rational curves (copies of the Riemann sphere $\mathbb{P}^1$), m_i positive integers (called *multiplicities*), and $S_u^2 = S_u \cdot S_u = 0$, since all fibers are homologous. The support of the divisor S_u is connected. We assume the fiber to be *minimal*, i.e. no Θ_i has self-intersection -1 (we can always reduce to this case by suitable blow-downs). The multiplicity of the fiber is $m \equiv \gcd(m_i)$. The fiber is *simple* iff $m = 1$: we focus on simple minimal fibers. The *type* of the simple fiber S_u is identified by the multiplicities $\{m_i\}$ and the way the various irreducible components Θ_i intersect each other. By general theory [39] the intersection form $\Theta_i \cdot \Theta_j \equiv -C_{ij}$ must be a negative semi-definite, symmetric, integral matrix, with non-negative off-diagonal entries and a zero-eigenvector with positive integral entries, $C_{ij}m_j = 0$. Indeed

$$0 \equiv \Theta_i \cdot S_u = \Theta_i \cdot S_u = \sum_j m_j \Theta_i \cdot \Theta_j = -C_{ij}m_j.$$

The adjunction formula [39] says that the diagonal entries are either 0 or -2 . It follows from Appendix 2 of Chap. 2 that C_{ij} is either the Cartan matrix of a simply-laced affine Kač–Moody Lie algebra or the zero matrix. C_{ij} is zero iff the fiber is irreducible $S_u = \Theta$. Otherwise $-\Theta_i \cdot \Theta_j$ is the Cartan matrix of an affine Lie algebra of type $\tilde{A}_{n-1}$, $\tilde{D}_{4+n}$, or $\tilde{E}_s$ ($s = 6, 7, 8$). In the irreducible case the fiber is a singular rational curve of virtual genus 1 (i.e. a singular cubic), and we have two possibilities: either the rational curve Θ has a simple node (Kodaira type **I**₁) or a cusp (Kodaira type **II**). The intersection matrices of types $\tilde{A}_1$ (resp. $\tilde{A}_2$) are realized by two different configurations of $\mathbb{P}^1$'s: either two (resp. three) rational curves tangent (resp. crossing transversely) at a single point which is Kodaira type **III** (resp. **IV**) or two (resp. three) rational curves crossing transversely at distinct points (types **I**₂ and **I**₃, respectively). All other $\tilde{A}\tilde{D}\tilde{E}$ intersection matrices are realized only by the configuration of rational curves dual to the affine Dynkin graph $\tilde{\Gamma}$ of the related Kač–Moody algebra: a simple root $\alpha_i \in \tilde{\Gamma}$ represents a rational component Θ_i , and two distinct components, Θ_i and Θ_j ($i \neq j$), cross transversely at a number of distinct points equal to $-C_{ij}$, each curve having intersection number -2 with itself. The equation $C_{ij}m_j = 0$ identifies the multiplicity m_i of Θ_i with the Coxeter label of the simple root α_i . The dual configurations to the Dynkin graphs of types $\tilde{A}_{b-1}$, $\tilde{D}_{4+b}$, $\tilde{E}_6$, $\tilde{E}_7$, and $\tilde{E}_8$ are called, respectively, exceptional fibers of Kodaira type **I**_b, **I**_b^{*}, **IV**^{*}, **III**^{*}, and **II**^{*}. Here $b \geq 0$: **I**₀ stands for a smooth elliptic curve, and **I**₁ for the rational curve with a simple node.

$$\phi \frac{dx}{y} \wedge dz_1 \wedge \cdots \wedge dz_m \quad (14.79)$$

where $z_1, \dots, z_m$ are holomorphic coordinates in B . Then ϕ , which by definition is a section of K_X , is a section of $K_B \otimes \mathcal{L}$, i.e. we have the *canonical bundle formula* for a Weierstrass elliptic fibration[34, 39]

$$K_X \simeq K_B \otimes \mathcal{L} \Rightarrow c_1(K_X) = -c_1(B) + c_1(\mathcal{L}), \quad (14.80)$$

and X is Calabi–Yau iff $c_1(\mathcal{L}) = c_1(B)$. Then, from Eq. (14.78),

$$c_1(\Delta) \equiv 12 c_1(\mathcal{L}) = 12 c_1(B). \quad (14.81)$$

In particular the anti-canonical divisor of B (if non-zero) is numerically effective.

Example: Elliptic K3's

Consider a K3 surface (CY 2-fold) K which is *elliptic*, i.e. is a holomorphic fibration over a curve B whose generic fibers are elliptic curves. K3 is simply connected, so the base B must be the Riemann sphere $\mathbb{P}^1$. Then $\mathcal{L} \rightarrow \mathbb{P}^1$ is a line bundle of degree $\deg \mathcal{L} = \chi(\mathbb{P}^1) \equiv 2$ and the discriminant D an effective divisor of degree 24. *Generically* D consists of 24 distinct points, so we have 24 singular fibers. The Euler characteristic of K3, 24, is the sum of the Euler characteristic of the singular fibers [37, 38, 46]. The generic elliptic K3 has 24 singular fibers of Euler characteristic 1, thus of type I_1 . The *generic* F-theory compactification on an elliptic K3 contains 24 7-branes which carry a unit of RR charge (the F-theory uplift of IIB $D7$'s). The most interesting case is when several of the points in D collide producing higher singular fibers and enhanced gauge groups.

★ **Tate Algorithm** One may read the type of the singular fibers (and hence the physics) directly from the equation (14.76) of the minimal Weierstrass model of the elliptic fibration. The method is called *Tate algorithm* [51–53]. First we distinguish the Kodaira fibers in two classes: *multiplicative* and *additive*. The distinction arises as follows: the set of smooth points of all fibers is an Abelian group of complex dimension 1. For smooth fibers the group is compact, hence an elliptic curve. For the singular fibers the group is non-compact and its connected component is either the multiplicative group $\mathbb{C}^\times$ or the additive group $\mathbb{C}$. In the first (resp. second) case we say that the singular fiber is multiplicative (resp. additive). The fibers of type I_n are multiplicative, and all others are additive. Let D_i be an irreducible component of the discriminant; we write $v(\Delta)$, $v(a)$ and $v(b)$ for, respectively, the order of vanishing along D_i of the discriminant Δ and of the coefficients a , b in the Weierstrass equation (14.76). One has [34]

$$\text{rank } \mathfrak{g} + 1 = \# \left(\begin{array}{c} \mathbb{P}^1 \text{'s in the} \\ \text{singular fiber} \end{array} \right) = \begin{cases} v(\Delta) & \text{multiplicative} \\ v(\Delta) - 1 & \text{additive} \end{cases} \quad (14.82)$$

Comparing with [37, 38, 46] we see that $v(\Delta)$ is also the Euler characteristic e of the fiber. Then the orders of vanishing of the Weierstrass coefficients are as in Table 14.3 (cf. Table 5.1 of [34]).

14.3.1 Duality Between M- and F-Theory

We claimed above that a static supersymmetric configuration with τ non-constant (and no higher form fluxes) requires X_{12} to be of the form $X_{12} = \mathbb{R}^{11-2m,1} \times CY_m$ ($m \leq 4$) where CY_m is a Calabi–Yau m -fold, in fact a special kind of CY which can be written as a holomorphic fibration $CY_m \rightarrow B_{m-1}$ whose *generic* fibers are elliptic curves. Such a Calabi–Yau is said to be *elliptic*. There are plenty of elliptic CYs: for a list of examples see [54]. This corresponds to F-theory with a 10d “gravitational brane” $\mathbb{R}^{11-2m,1} \times B_{m-1}$.

Table 14.3 Vanishing orders of coefficients of Weierstrass equation for the several Kodaira fibers

	$v(\Delta) \equiv e$		$v(a), v(b)$		or	$v(a), v(b)$		
I_0	0		$0, \geq 0$		or	$\geq 0, 0$		
I_0^*	6		$2, \geq 3$		or	$\geq 2, 3$		
	$I_{n>0}$	II	III	IV	$I_{n>0}^*$	IV^*	III^*	II^*
$v(\Delta) \equiv e$	n	2	3	4	$6+n$	8	9	10
$v(a)$	0	≥ 1	1	≥ 2	2	≥ 3	3	≥ 4
$v(b)$	0	1	≥ 2	2	3	4	≥ 5	5

To show the claim, we consider M-theory on the 11d manifold $\mathbb{R}^{10-2m,1} \times CY_m$ with CY_m a *strict* Calabi–Yau m -fold which is elliptic over the complex base $B \equiv B_{m-1}$ ($1 \leq m \leq 4$). This produces an effective theory in $(11 - 2m)$ -dimensions which preserves 2^{6-m} supercharges. The volume v of the elliptic fibers over a point $p \in B$ is independent of the point²⁰ and by Yau’s Theorem 11.9 we may find a CY metric on CY_m with fibers of volume v as small as we wish. Up to terms of higher order in v , the M-theory metric takes the form [55]

$$ds_{11}^2 = \eta_{\mu\nu} dX^\mu dX^\nu + ds_{CY}^2 = \eta_{\mu\nu} dX^\mu dX^\nu + ds_B^2 + \frac{v}{\tau_2} \left((dx + \tau_1 dy)^2 + \tau_2^2 dy^2 \right) \quad (14.83)$$

where

- X^μ ’s are the coordinates of the flat factor $\mathbb{R}^{10-2m,1}$;
- x, y are periodic coordinates on the elliptic fiber both of period 1;
- ds_B^2 is a Kähler metric on the base B ;
- τ_1 , and τ_2 are functions on B .

We call a -cycle (resp. b -cycle) the cycle in the elliptic fiber which is dual to dx (resp. dy). Making the a -cycle small, we get weakly-coupled Type IIA on a 10d space $\mathbb{R}^{10-2m,1} \times Y$ where Y is a fibration over B with S^1 fibers of coordinate $y \sim y + 1$. We T -dualize the S^1 fiber getting Type IIB on a dual space $\mathbb{R}^{10-2m,1} \times \tilde{Y}$ with $\tilde{Y} \rightarrow B$ the dual fibration.²¹

We rewrite the 11d metric in terms of the Type IIA metric in the string frame²²

²⁰ Indeed, v is the Kähler form ω of CY_m integrated on the fiber, and is the same for all fibers since ω is closed and the fibers are all homologous.

²¹ T -dual fibrations of general S^1 principal bundles $Y \rightarrow B$ are defined and studied in Sect. 6.4.2.

²² To get Eq. (14.84) we used Eqs. (8.33) and (8.44). We also shifted the IIA dilaton Φ_{IIA} to make the small size v to appear explicitly in the formula.

$$ds_{11}^2 = v e^{4\Phi_{\text{IIA}}/3} (dx + C^{(1)})^2 + e^{-2\Phi_{\text{IIA}}/3} ds_{\text{IIA}}^2 \quad (14.84)$$

where $C^{(1)}$ is the RR gauge vector. Comparing with (14.83) we get the IIA quantities

$$\begin{aligned} C^{(1)} &= \tau_1 dy, & e^{4\Phi_{\text{IIA}}/3} &= \frac{1}{\tau_2}, & g_{\text{IIA}} &= \frac{e^{\Phi_{\text{IIA}}}}{(M_{11}\sqrt{\alpha'})^3}, \\ R_9 &= \sqrt{v} e^{\Phi_{\text{IIA}}} \tau_2, & \frac{1}{\alpha' M_{11}^3} &= \sqrt{v}, & ds_{\text{IIA}}^2 &= \frac{1}{\sqrt{\tau_2}} (v \tau_2 dy^2 + ds_9^2), \end{aligned} \quad (14.85)$$

where

$$ds_9^2 \equiv \eta_{\mu\nu} dX^\mu dX^\nu + ds_B^2. \quad (14.86)$$

Taking the T -duality along the y -circle, and using the Buscher rules (Sect. 6.4.2), we get Type IIB with $C^{(0)} \equiv (C^{(1)})_y = \tau_1$ and

$$e^{\Phi_{\text{IIB}}} \equiv g_{\text{IIB}} \equiv \frac{\sqrt{\alpha'}}{R_9} g_{\text{IIA}} = \frac{e^{\Phi_{\text{IIA}}}}{(M_{11}^3 \alpha')(\sqrt{v} e^{\Phi_{\text{IIA}}} \tau_2)} = \frac{1}{\tau_2} \quad (14.87)$$

so that the Type IIB complex scalar

$$C^{(0)} + \frac{i}{g_{\text{IIB}}} \equiv \tau_1 + i \tau_2 \equiv \tau, \quad (14.88)$$

is nothing else than the modulus τ of the elliptic fiber of the Calabi–Yau $CY_m \rightarrow B$. The string frame IIB metric is

$$ds_{\text{IIB}}^2 = \frac{1}{\sqrt{\tau_2}} \left(\frac{dy^2}{M_{11}^6 v^2} + ds_9^2 \right) \equiv \frac{1}{\sqrt{\tau_2}} \left(\frac{(\alpha')^2}{v} dy^2 + ds_9^2 \right). \quad (14.89)$$

The coordinate $Y \equiv \alpha' y / \sqrt{v}$ is now periodic of period $\alpha' / \sqrt{v}$. The Einstein frame IIB metric is (cf. Eq. (8.76))

$$(ds_{\text{IIB}}^2)_{\text{Einstein}} = ds_9^2 + dY^2 \equiv \eta_{\mu\nu} dX^\mu dX^\nu + dY^2 + ds_B^2. \quad (14.90)$$

In the limit of zero fiber volume, $v \rightarrow 0$, the coordinate Y decompactifies, and the IIB spacetime becomes $\mathbb{R}^{11-2m,1} \times B$. Over the spacetime we have an elliptic fibration

$$X_{12} \equiv \mathbb{R}^{11-2m,1} \times CY_m \xrightarrow{\text{Id} \times \pi} \mathbb{R}^{11-2m,1} \times B_{m-1} \quad (14.91)$$

where the fiber over p has period

$$\tau(p) = C^{(0)}(p) + i e^{-\Phi_{\text{IIB}}(p)} \quad (14.92)$$

as claimed. Note that in the limit $v \rightarrow 0$, the “new” non-compact coordinate Y gets unified with the original non-compact coordinates X^μ in the sense that the full $(12 - 2m)$ -dimensional Poincaré symmetry gets “magically” restored.

14.4 Matrix Theory: A Proposal for M-Theory

We have inferred the existence of M-theory from the strong coupling limit of the IIA superstring. At the non-perturbative level M-theory looks as *the* fundamental theory, while the various superstring theories are “merely” perturbative asymptotic expansions at certain corners of its moduli space (cf. Fig. 13.2). It would be highly desirable to have a direct, first-principle, definition of the fundamental M-theory. In this section we sketch a nice proposal called *Matrix Theory* [56–59].

The first step in any formulation of M-theory is to identify its fundamental degrees of freedom. In view of the opening remarks in Sect. 13.5, it is natural to think that they are D0-branes. Let us clarify the physical reasoning beyond this idea [60, 61]. We start with an arbitrary *normalizable* state of the IIA superstring and we “boost” it to a large momentum along the M-theory circle. Here the notion of “boosting” is slightly imprecise, since the compactification on the M-theory circle breaks the 11d Lorentz invariance; however at large coupling (hence large R_{10}) the Lorentz symmetry is approximately restored, and “boosting” makes sense asymptotically. The energy of a state with $n \gg M_{11} R_{10} \gg 1$ units of compact momentum is

$$E = (p_{10}^2 + \mathbf{q}^2 + m^2)^{1/2} \approx p_{10} + \frac{\mathbf{q}^2 + m^2}{2p_{10}} = \frac{n}{R_{10}} + \frac{R_{10}}{2n}(\mathbf{q}^2 + m^2). \quad (14.93)$$

Here $\mathbf{q}$ is the momentum in the non-compact 9 spatial dimensions and m is the total mass of the state. Recalling the relation between p_{10} and the D0-charge, Sect. 13.5, the energy of this “boosted” state satisfies a BPS bound

$$E \geq \frac{n}{R_{10}} \quad (14.94)$$

which is saturated iff the state is a BPS bound state of n D0’s. States which have finite energy in the original frame have

$$E - n/R_{10} = O(R_{10}/n) \quad (14.95)$$

in the boosted frame. That is, all finite-energy states become “almost BPS” after the large boost. They may be seen as a “composed state” of a large number n of D0’s and other objects. We can describe such states from the viewpoint of the D0’s world-line theory whose d.o.f. arise from open strings attached to the D0-branes. Only *low-energy* open string states are relevant in the limit. The low-energy effective

Hamiltonian for the SQM on the DO's world line was given in Eq. (12.61). In terms of the M-theory parameters M_{11} and R_{10} it reads

$$H = R_{10} \operatorname{tr} \left[\frac{1}{2} p_i p_i - \frac{M_{11}^6}{16\pi^2} [X^i, X^j]^2 - \frac{M_{11}^3}{4\pi} \lambda \Gamma^0 \Gamma^i [X^i, \lambda] \right]. \quad (14.96)$$

The higher order terms in the Born–Infeld Hamiltonian have larger powers of momentum, and hence are suppressed by the large boost. We subtracted the constant n/R_{10} from the Hamiltonian. The Hamiltonian (14.96) is conjectured [56] to give the full description of the systems with

$$p_{10} = n/R_{10} \gg M_{11}^{-1}. \quad (14.97)$$

Now take R_{10} and n/R_{10} to infinity, to describe a highly boosted system in eleven non-compact dimensions. By 11-dimensional Lorentz invariance, we can put any system in this frame, so this should be a complete description of the whole M-theory! This is the *Matrix Theory* proposal for the fundamental M-theory. This is a conjecture, not a derivation: the Hamiltonian (14.96) was identified in the IIA corner of the moduli space in Fig. 13.2 and we are declaring it to be valid everywhere.

This looks like an amazingly simple and explicit formulation: we have a supersymmetric Quantum Mechanics whose bosonic d.o.f. are the $n \times n$ matrices X_{ab}^i ($a, b = 1, \dots, n, i = 1, \dots, 9$). The SQM is super-Yang–Mills in 1d with 16 supercharges and gauge group $U(n)$. The $U(1)$ d.o.f. correspond to the center-of-mass motion and decouple, leaving a SYM with $G = SU(n)$. However its simplicity is merely apparent: the description becomes valid only in the limit $n \rightarrow \infty$, so we have a strongly interacting quantum system with an *infinite* number of degrees of freedom. A system with infinitely many d.o.f. contains states of arbitrary large complexity, as the examples below will illustrate.

The description of the 11-dimensional spacetime in Matrix theory looks very asymmetric: time t is an explicit coordinate, nine spatial coordinate emerges from the matrix quantum operators X^i , and the last coordinate arises as the decompactification limit of the angle θ which is Fourier-dual to the quantized momentum $n \in \mathbb{Z}$ which is the size of the matrix-operators X^i .

Some Checks that the Proposal is on the Right Track

Matrix theory, if correct, is a consistent quantum theory of gravity, hence should not contain any adjustable constant and only one scale, namely the 11d Planck mass

$$M_{11} = (g^{1/3} \alpha'^{1/2})^{-1}. \quad (14.98)$$

At first sight the Hamiltonian (14.96) seems to involve the parameter R_{10} . Recall, however, that the system is boosted and so internal times τ get dilated. The boost factor is proportional to p_{10} , hence the time scale should be multiplied by the dimensionless factor

$$\frac{p_{10}}{M_{11}} = \frac{n}{M_{11} R_{10}} \quad (14.99)$$

and only the scale M_{11} appears in the quantum evolution operator $\exp(-iHt)$.

Gravitational Physics from Matrices As in the IIA/M-theory duality, a graviton of momentum $p_{10} = n/R_{10}$ is a bound state of n D0-branes. The existence of these bound states is necessary for Matrix theory to be correct. To check their presence we use an argument akin to the non-perturbative study of SUSY bound states in Sect. 12.7.1. The 1d $U(n)$ gauge theory decouples into an Abelian $U(1)$ sector and a non-Abelian $SU(n)$ one. The $SU(n)$ dynamics yields a *gapped* zero-energy bound state (a SUSY vacuum) as in Sect. 12.7.1, and only the Abelian d.o.f. remain visible at low-energy. For a bound state of total momentum $q_i \equiv \text{tr}(p_i)$, the conjugate momenta $n \times n$ matrices p_i then take the form $p_i = q_i \mathbf{1}_n/n$, and the energy (14.96) becomes

$$E = \frac{R_{10}}{2} \text{tr}(p_i p_i) = \frac{\mathbf{q}^2}{2p_{10}}, \quad (14.100)$$

reproducing the correct energy (14.93) for a massless state.

Next we consider an elementary physical process: say graviton-graviton scattering. Let the two gravitons have momenta in the compact 10-th direction

$$p_{10} = n_{1,2}/R_{10} \quad (14.101)$$

and be at well-separated positions Y_1^i and Y_2^i in the transverse directions. The total number of D0's is $n_1 + n_2$, and the coordinate matrices X^i are approximately block diagonal. We write

$$X^i = X_0^i + x^i \equiv \overbrace{\begin{pmatrix} Y_1^i \mathbf{1}_{n_1} & 0 \\ 0 & Y_2^i \mathbf{1}_{n_2} \end{pmatrix}}^{\text{"classical position"}} + \overbrace{\begin{pmatrix} x_{11}^i & x_{12}^i \\ x_{21}^i & x_{22}^i \end{pmatrix}}^{\text{"quantum fluctuation"}}. \quad (14.102)$$

The off-diagonal d.o.f. x_{12}^i and x_{21}^i gets very heavy for large separations $|Y_1^i - Y_2^i| \gg M_{11}^{-1}$: the commutator

$$[X_0^i, x_{12}^j] = (Y_1^i - Y_2^i) x_{12}^j \quad (14.103)$$

gives them a mass proportional to the separation of the gravitons. Thus we can integrate out the off-diagonal d.o.f. to obtain the effective interaction between the two gravitons. In the limit $|Y_1^i - Y_2^i| \rightarrow \infty$ the off-diagonal blocks x_{12}^i, x_{21}^i go to zero, and the system decouples into a $SU(n_1)$ and a $SU(n_2)$ 1d SYM since $[x_{11}^i, x_{22}^j] = 0$. In this limit, the wave-function is the product of the two bound-state wave-functions

$$\psi(x_{11}, x_{22}) \approx \psi_0(x_{11}) \psi_0(x_{22}). \quad (14.104)$$

We would like to use this process to test the Matrix theory proposal: if correct the effective interaction at long distance produced by integrating out the off-diagonal

blocks x_{12}^i, x_{21}^i should reproduce the graviton-graviton force of 11d supergravity. Luckily, this check does not require any computation: the necessary expressions can be read from the cylinder amplitude (12.133). At distance small compared to the string scale, the cylinder is dominated by the lightest open strings stretched between the D0-branes, which are precisely the off-diagonal Matrix theory degrees of freedom. At distances large compared with the string scale the cylinder is dominated by the lightest closed string states and so goes over to the supergravity result. This is 10d SUGRA, but it is equivalent to the answer from the 11-dimensional supergravity for the following reason. In the process we are studying the sizes of the blocks stay fixed at n_1 and n_2 meaning that the values of p_{10} and p'_{10} do not change in the scattering and the p_{10} of the exchanged graviton is zero. This has the effect of averaging over x^{10} and so giving the dimensionally reduced answer.

Finally, we should keep only the leading velocity dependence from the cylinder, because of the time dilation from the boost suppresses higher powers as in Eq. (14.93). Multiplying by the number of D0-branes in each bunch our previous result (12.137) (with $p = 0$), we get

$$V(r, v) = 4\pi^{5/2} \Gamma(7/2) \alpha'^3 n_1 n_2 \frac{v^4}{r^7} = \frac{15\pi^3}{2} \frac{p_{10} p'_{10}}{M_{11}^9 R_{10}} \frac{v^4}{r^7}. \quad (14.105)$$

Because the functional form is the same at large and small r , the matrix theory correctly reproduces the supergravity amplitude.

14.4.1 The M-Theory Membrane

To be correct Matrix theory should contain states which represent all objects of M-theory in particular the extended objects such as M2- and M5-branes and more general multi-brane configurations. Strikingly, Matrix theory does incorporate all kinds of M-theory brane arrangements (and a lot more) [62, 63]! Here we limit ourselves to show how the M2-branes emerge from the D0-brane Hamiltonian.²³ Arguably the M2 membrane is the most fundamental object in M-theory which encodes all its fundamental d.o.f. much as the F1 string does in superstring theory: in fact, as we know from Chap. 13, the F1 of Type IIA arises a special limit of the wrapped M2 brane.

To see how the M2 arises in Matrix theory we first need to quantize the super 2-brane in the light-cone gauge and then compare the result with the D0-brane SQM (14.96) in the Hilbert-space sector describing a single 2-brane. For the first item see Box 14.7. For the second issue we need some algebraic preliminary.

²³ Naively one may have expected that to describe M2-branes one needs to add D2-brane d.o.f. to the D0 ones. This would kill the Matrix theory as a prospective *complete* description of M-theory. But “miraculously” M2 branes are automatically present in the Matrix SQM as collective excitations of the D0 d.o.f. and Matrix theory passes also this highly non-trivial check.

BOX 14.7 Super-membrane in light-cone gauge

We quantize the supermembrane moving in flat 11 dimensions following [64]

Index convention: $\mu, \nu = 0, \dots, 10; i, j = 0, 1, 2; a, b = 2, \dots, 10; r, s = 1, 2$.

Super-Nambu–Goto Lagrangian: $\mathcal{L} = -\sqrt{-\det g} + \text{fermions}, \quad g_{ij} \equiv \partial_i X^\mu \partial_j X^\nu \eta_{\mu\nu}.$

Light-cone gauge: $y_0 = X^+$. One has

$$\begin{aligned} G_{rs} &\stackrel{\text{def}}{=} g_{rs} = \partial_r X^a \partial_s X^a, & u_r &\stackrel{\text{def}}{=} g_{0r} = \partial_r X^- + \partial_0 X^a \partial_r X^a, \\ g_{00} &= 2\partial_0 X^- + \partial_0 X^a \partial_0 X^a & -\det g &= \Delta G, \\ \Delta &\stackrel{\text{def}}{=} -g_{00} + u_r G^{rs} u_s, & G &\stackrel{\text{def}}{=} \det G_{rs} \end{aligned}$$

and the gauge-fixed Lagrangian is simply

$$\mathcal{L} = -\sqrt{\Delta G} + \text{fermions}.$$

The conjugate momenta are

$$\begin{aligned} P^a &= \frac{\partial \mathcal{L}}{\partial(\partial_0 X^a)} = \sqrt{\frac{G}{\Delta}} \left(\partial_0 X^a - u_r G^{rs} \partial_s X^a \right) + \text{fermions}, \\ P^+ &= \frac{\partial \mathcal{L}}{\partial(\partial_0 X^-)} = \sqrt{\frac{G}{\Delta}} + \text{fermions} \end{aligned}$$

so that

$$\textbf{Hamiltonian :} \quad H = P^a \partial_0 X^a + P^+ \partial_0 X^- - \mathcal{L} = \frac{P^a P^a + G}{2P^+} + \text{fermions}$$

$$\text{with} \quad G = \det G_{rs} = \frac{1}{2} \epsilon^{rt} \epsilon^{su} \partial_r X^a \partial_s X^a \partial_t X^b \partial_u X^b = \frac{1}{2} \{X^a, X^b\} \{X^a, X^b\}$$

$$\textbf{where} \quad \{X^a, X^b\} \stackrel{\text{def}}{=} \partial_1 X^a \partial_2 X^b - \partial_2 X^a \partial_1 X^b \equiv \textbf{Poisson bracket}$$

The residual gauge symmetries are the time-independent, area preserving diffeomorphisms ($\equiv$ symplectomorphisms [65] $\equiv$ canonical transformations [66]) of the two-dimensional slices at constant y^0 . The full Hamiltonian reads [64]

$$H = \frac{P^a P^a}{2P^+} + \frac{1}{4P^+} \{X^a, X^b\}^2 - \theta \gamma_- \gamma_a \{X^i, \theta\} \quad (\clubsuit)$$

where we reinserted the fermions θ which are space-time spinors. Here X^a and θ are the transverse target superspace coordinates.

Quantum Torus Algebra, Weyl Algebra, and the Moyal Product *

We define the $n \times n$ matrices

$$U = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \alpha & 0 & 0 & \cdots \\ 0 & 0 & \alpha^2 & 0 \\ 0 & 0 & 0 & \alpha^3 \\ \vdots & & & \ddots \end{pmatrix} \quad V = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}, \quad (14.106)$$

where $\alpha \equiv \exp(2\pi i/n)$. In the index notation the matrices U and V read

$$U_{kl} = e^{2\pi i k/n} \delta_{kl} \quad \text{and} \quad V_{kl} = \delta_{k, l+1 \bmod n}. \quad (14.107)$$

V is known as the *circulant* matrix. These matrices satisfy the relations

$$U^n = V^n = 1, \quad UV = \alpha VU, \quad \alpha^n = 1, \quad (14.108)$$

that is, they yield a representation of the *quantum torus algebra* [67, 68] with α a root of unity (see Box 14.8 for background). The quantum torus algebra is strictly related to the Weyl algebra in Quantum Mechanics with Planck constant $\hbar = 1/n$ (see Box 14.8). The quantum torus algebra with α a root of 1 is Morita equivalent to the classical torus algebra (i.e. to the algebra of functions $S^1 \times S^1 \rightarrow \mathbb{C}$) [67], and U, V correspond to the trivial representation of the classical algebra: therefore the relations (14.108) determine the matrices U and V up to conjugacy in $GL(n, \mathbb{C})$. By the very meaning of Morita equivalence for associative algebras [69], the algebra generated over $\mathbb{C}$ by U and V is isomorphic to the algebra of $n \times n$ complex matrices. In particular, by the Poincaré–Birkhoff theorem [70] the n^2 matrices

$$U^r V^s \quad 0 \leq r, s \leq n-1 \quad (14.109)$$

form a basis of the space of all $n \times n$ matrices. Given a $n \times n$ matrix M , we can expand it in this basis

$$M = \sum_{r,s=0}^{n-1} M_{r,s} U^r V^s \quad (14.110)$$

and then associate to the matrix M a periodic function of two discrete variables, namely the restriction of the *symbol* of the corresponding quantum operator acting on $L^2(\mathbb{R})$ (cf. Box 14.8)

$$M(p, q) \stackrel{\text{def}}{=} \sum_{r,s=0}^{n-1} M_{r,s} \exp(2\pi i p r + i q s) \quad (14.111)$$

to the discrete set of points

$$p = k/n \quad \text{and} \quad q = 2\pi l/n \quad \text{where } k, l \in \mathbb{Z}. \quad (14.112)$$

In the limit $n \rightarrow \infty$ the discrete quantities p and q become continuous real variables. The original matrix M and its restricted symbol $M(k/n, 2\pi l/n)$ give two equivalent ways of writing the $n \times n$ matrix algebra: the matrix product becomes (the discrete restriction of) the *Moyal product* $*$ of symbols [71–74]²⁴

$$\begin{aligned} MN(p, q) &= M(p, q) * N(p, q) \stackrel{\text{def}}{=} \\ &\stackrel{\text{def}}{=} \lim_{\substack{p' \rightarrow p \\ q' \rightarrow q}} \exp \left[-\frac{i}{n} \frac{\partial}{\partial q} \frac{\partial}{\partial p'} \right] M(p, q) N(p', q'). \end{aligned} \quad (14.113)$$

The Moyal algebra of symbols coincides with the deformation quantization of the algebra of classical functions on the 2-torus seen as a phase space with symplectic structure $dq \wedge dp$ [75]. The matrix trace is given by the integral over the phase space divided by the volume of the phase-space quantum cell²⁵ $2\pi \hbar \equiv 2\pi/n$

$$\text{Tr}(M) = \sum_{r,s=0}^{n-1} M_{r,s} \text{Tr}(U^r V^s) = n M_{0,0} = \frac{n}{2\pi} \int_0^1 dp \int_0^{2\pi} dq M(p, q). \quad (14.114)$$

M2 Branes in Matrix Theory

The degrees of freedom of the SQM on the world-line of n D0's, X^i and λ_α , are $n \times n$ matrices with operator entries. Hence we can see them as operator-valued elements in the quantum torus algebra $\mathbb{T}(e^{2\pi i/n})$ and consider their symbols $X^i(p, q)$, $\lambda_\alpha(p, q)$ as in the previous paragraph. In the Matrix theory limit $n \rightarrow \infty$ these symbols become *operator-valued functions* on the classical phase space $(S^1)^2$ with periodic coordinates $p \sim p + 1$ and $q \sim q + 2\pi$. For a general M-theory state the dependence of these operators on (p, q) is expected to be quite quirky. However there are also “regular” quantum states which produce (approximatively) smooth symbols. In this case the Moyal product $*$ has a regular limit, and we can easily compute the symbols of the commutators

$$\begin{aligned} [X^i, X^j](p, q) &\equiv X^i(p, q) * X^j(p, q) - X^j(p, q) * X^i(p, q) = \\ &= \frac{i}{n} (\partial_q X^j(p, q) \partial_p X^i(p, q) - \partial_p X^j(p, q) \partial_q X^i(p, q)) + O\left(\frac{1}{n^2}\right) \equiv \\ &\equiv \frac{i}{n} \{X^j(p, q), X^i(p, q)\}_{\text{PB}} + O\left(\frac{1}{n^2}\right), \end{aligned} \quad (14.115)$$

where $\{\cdot, \cdot\}_{\text{PB}}$ stands for the Poisson bracket on the phase space with symplectic form $dq \wedge dp$. Equation (14.115) is just the basic fact in deformation quantization: to leading order as $\hbar \equiv \frac{1}{n} \rightarrow 0$, the commutator is given by $i\hbar$ times the Poisson bracket [75]. We stress that this limit is well-defined for a “small” class of “regular” states of

²⁴ Here the derivatives are defined algebraically, e.g. $\partial_q M(p, q) \equiv i \sum_{r,s=0}^{n-1} s M_{r,s} e^{2\pi i p r + i q s}$, and make sense even when p and q are *discrete* variables.

²⁵ This is a manifestation of the Heisenberg indetermination principle [76].

BOX 14.8 Quantum torus algebra and all that

Let $q \equiv e^{2\pi i \tau} \in \mathbb{C}^\times$ (we can assume $|q| \leq 1$ without loss). The *quantum torus algebra* [67] (a.k.a. the *quantum plane* [68]) $\mathbb{T}(q)$ is the algebra generated over $\mathbb{C}$ by two *invertible* elements U and V which satisfy the commutation relation

$$U V = q V U. \text{quad } (\ddagger).$$

For $q = 1$ ($\tau \in \mathbb{Z}$) this is just the classical polynomial algebra in two indeterminates (i.e. the coordinate ring of the affine plane); for $q \neq 1$ we get a quantization of this algebra.

Let p and q canonical quantum operators acting on $L^2(\mathbb{R})$ with $[q, p] = i\hbar \equiv i\tau$. We consider the Weyl operators

$$U = \exp(2\pi i p), \quad V = \exp(i q) \quad (\clubsuit).$$

Clearly they satisfy Eq. $(\ddagger)$ so that the Hilbert space $L^2(\mathbb{R})$ carries a representation of $\mathbb{T}(q)$ for all q (unitary iff $|q| = 1$) given by

$$\varrho := U \mapsto U, \quad V \mapsto V.$$

This representation is highly reducible since its centralizer is the dual quantum torus algebra generated by $\tilde{U} \tilde{V}$ [77]

$$\tilde{U} \tilde{V} = \tilde{q} \tilde{V} \tilde{U}, \quad \text{where } \tilde{q} = e^{-2\pi i / \tau}, \quad \tilde{U} = \exp(2\pi i p / \tau), \quad \tilde{V} = \exp(-i q / \tau).$$

The case $\tau \in \mathbb{Q}$ When $\tau \in \mathbb{R}$ we may endow $\mathbb{T}(\zeta)$ with a C^* -algebra structure by setting $U^\dagger U^{-1}$ and $V^\dagger = V^{-1}$. If $q \equiv \zeta$ is a n -th root of unity, the C^* -algebra $\mathbb{T}(\zeta)$ has a large center $Z(\mathbb{T}(\zeta))$ generated by the invertible elements U^n and V^n . $Z(\mathbb{T}(\zeta))$ is a commutative $*$ -algebra isomorphic to the $*$ -algebra of bounded functions on the torus $S^1 \times S^1$ through the map

$$f(\theta, \phi) \equiv \sum_{r,s} f_{r,s} e^{ir\theta + is\phi} \rightsquigarrow \sum_{r,s} f_{r,s} U^{rn} V^{sn}.$$

Hence the quantum torus algebra at a n -th root of unity is isomorphic to the algebra of $n \times n$ matrices whose entries are bounded functions $S^1 \times S^1 \rightarrow \mathbb{C}$ through

$$U \rightarrow e^{i\theta/n} U, \quad V \rightarrow e^{i\phi/n} V, \quad 0 \leq \theta, \phi \leq 2\pi n,$$

where U, V are the matrices in Eq. (14.106).

the original matrix SQM system. It is easy to write down the quantum Hamiltonian (14.96) in the strict limit $n \rightarrow \infty$ for such “regular” states using Eqs. (14.114) and (14.115). In this “regular” Hilbert-space sector, the limit is smooth and one gets

$$R_{10} \int dp dq \left(\frac{n}{2} \Pi_i \Pi_i + \frac{M_{11}^6}{16\pi^2 n} \{X^i, X^j\}_{\text{PB}}^2 - i \frac{M_{11}^3}{4\pi} \lambda \Gamma^0 \Gamma^i \{X^i, \lambda\}_{\text{PB}} \right). \quad (14.116)$$

In this “regular” sector, the $n = \infty$ Hamiltonian is *identical* to the Hamiltonian of the 11-dimensional supersymmetric membrane in the light-cone gauge (see Box 14.7). This shows that M2 states are present in the Matrix theory Hilbert space.

We illustrate this conclusion in a simple example.

Example: A static toroidal M2 brane

Consider the configuration

$$X^1(p, q) = a q, \quad X^2(p, q) = b p \quad (14.117)$$

with a, b constants. Since q and p are periodic variables, this is consistent if the coordinates X^1, X^2 are also periodic, i.e. the 11d space-time has the form $(S^1)^2 \times M$ and the membrane is wrapped on the 2-torus factor. The energy of this configuration is

$$\frac{M_{11}^6 R_{10} a^2 b^2}{8\pi^2 n} = \frac{M_{11}^6 A^2}{2(2\pi)^4 p_{10}} = \frac{\tau_{M2}^2 A^2}{2 p_{10}} \quad (14.118)$$

where $A \equiv 2\pi ab$ is the area of the M2. The product $\tau_{M2} A$ is the mass of a M2-brane of this area, so it agrees with the energy (14.93) for $\mathbf{q} = 0$ (recall that the constant term n/R_{10} was subtracted).

Multi-Brane Configurations

The SQM system (14.96) with $n = \infty$ should describe *all* states of M-theory. In particular its Hilbert space should contain sectors which represent the states of general configurations of any number N of M2-branes. This is indeed the case: the argument is similar to the one we used before to show that Matrix theory describes states with two gravitons (hence with any number of them). Consider a configuration where the X^i have a common block structure in the sense that there is a partition of n of the form $n = \sum_{s=1}^N m_s$ and

$$X^i \approx \begin{pmatrix} X_1^i & & & \\ & X_2^i & & \\ & & \ddots & \\ & & & X_N^i \end{pmatrix}, \quad (14.119)$$

where X_s^i are $m_s \times m_s$ matrix blocks, and the off-diagonal blocks are supposed to be small, that is, heavy: this requires the eigenvalues of $X_s^i - X_t^i$ to be all large for $t \neq s$. As before we introduce symbols for each block-matrix

$$X_s^i \rightsquigarrow X_s^i(p_s, q_s) \quad p_s = k_s/m_s, \quad q_s = 2\pi l_2/m_s. \quad (14.120)$$

Suppose that all symbols are regular as $m_s \rightarrow \infty$ ($s = 1, \dots, N$). Neglecting the off-diagonal blocks, the limiting Hamiltonian is the sum of N non-interacting copies of the light-cone supermembrane Hamiltonian (14.116). The limiting “regular” wave-eigenfunctions clearly describe states with N M2 branes. The off-diagonal blocks can be neglected only for large $X_s^i - X_t^i$, i.e. when the branes are far away from each other. For finite non-zero separations, integrating away the massive off-diagonal d.o.f. produces interactions between branes. (These interactions vanish whenever the multi-brane state is BPS).

14.5 6d (2, 0) SCFTs

General references for SCFT in six-dimensions are [78–93].

We have seen in Sect. 13.6 that the world-volume effective theory on a stack of N Type IIA NS5-branes — that now we interpret non-perturbatively as N M-theory M5-branes — is a “non-Abelian” 6d SCFT with chiral $(2, 0)$ supersymmetry. When compactified on a circle we get a 5d theory, which at low-energy looks like 5d super-Yang–Mills with gauge group $G = U(N)$ and 16 supercharges, i.e. the effective theory on a stack of D4’s. The actual 5d theory is a UV completion of 5d SYM which is expected to be the 6d theory itself.

When $N = 1$ the 6d SCFT consists of just one free $(2, 0)$ tensor supermultiplet which contains a *chiral* gauge 2-form $B_{\mu\nu}^+$ with self-dual field-strength 3-form, 5 real scalars χ^a in the vector of $\mathfrak{so}(5) \simeq \mathfrak{sp}(2)$, and two symplectic-Majorana–Weyl fermions of the same chirality in the $\mathbf{4}$ of the R-symmetry $Sp(2)$ (see Box 13.3). The v.e.v. of the scalars χ^a may be thought of as the transverse positions of the M5 brane in 11 dimensions. If we have several M5’s which are well-separated in the transverse directions, the light scalars correspond to the positions of the several branes and the world-volume theory is approximated by N copies of the free theory. When the distances between the M5’s get small, we get new light degrees of freedom from the M2’s suspended between them: these are tensionless strings which yield the “non-Abelian enhancement” of the theory. The overall $U(1)$ part corresponds to the center-of-mass motion of the brane system, which is a free tensor supermultiplet. We remain with an interacting 6d theory which after compactification to 5d reduces the IR to $SU(N)$ SYM. However the group $SU(N)$ is not a symmetry of the 6d system, and the model is rather labeled by the corresponding Lie algebra written in Cartan’s notation: we use the symbol A_{N-1} for the interacting sector of the theory living on a stack of N parallel M5’s.

That the non-Abelian d.o.f. are strings is evident from the fact that in the tensor multiplets we have 2-form gauge fields whose source must be a string. The field-strength is self-dual, and there is no distinction between electric and magnetic sources: *the string is self-dual*.

The Tensor Branch

Just as in the case of $(1, 1)$ the v.e.v. of the vectors’ scalars parametrize the Coulomb branch (cf. Box 12.2), in a $(2, 0)$ SCFT the v.e.v. of the tensors’ scalars parametrize the *tensor branch* $\mathcal{T}$. The two branches look very similar, since the 6d tensor branch gets identified with the 5d SYM Coulomb branch by compactification on a large circle. Away from points where we have tensionless strings, the tensor branch parametrizes the transverse positions of N indistinguishable branes so that, neglecting the decoupled center of mass position, we expect

$$\text{tensor branch of 6d } (2, 0) \text{ SCFT of type } A_{N-1} = (\mathbb{R}^5)^{N-1} / \mathfrak{S}_N \quad (14.121)$$

where the symmetric group $\mathfrak{S}_N$ acts through its irreducible $(N - 1)$ -dimensional representation. The RHS may be written in a more illuminating way. Let $\mathfrak{h}_{A_{N-1}} \subset A_{N-1}$ be the Cartan subalgebra and $\text{Weyl}(A_{N-1})$ the Weyl group of $A_{N-1} \equiv \mathfrak{su}(N)$

$$\text{tensor branch} = (\mathfrak{h}_{A_{N-1}} \times \mathfrak{h}_{A_{N-1}} \times \mathfrak{h}_{A_{N-1}} \times \mathfrak{h}_{A_{N-1}} \times \mathfrak{h}_{A_{N-1}}) / \text{Weyl}(A_{N-1}) \quad (14.122)$$

where the Weyl group acts diagonally. Of course this is just the Coulomb branch in 5d of maximally supersymmetric Yang–Mills theory with gauge group $SU(N)$.

The tensor branch geometry (14.122) suggests that there must be other interacting 6d (2, 0) SCFTs whose compactification to 5d lead to (the UV completion of) maximal supersymmetric Yang–Mills with an arbitrary simple gauge group G . Indeed, it looks implausible that 5d maximal SYM admits a UV completion for $G = SU(N)$ but not for other gauge groups. It is also plausible that when the UV completion exists, it is unique. Indeed the existence of UV-complete interacting QFTs in 5 or more dimensions is already an overdetermined “unlikely miracle” and there cannot be too many of them.²⁶ In particular there are no non-free weakly coupled QFT in $d \geq 5$, so our SCFTs should be inherently strongly coupled (and they cannot contain any adjustable parameter).

The Coulomb branch of maximal SYM in 5d is

$$(\text{Cartan algebra of } \mathfrak{g})^5 / (\text{Weyl group of } \mathfrak{g}), \quad \mathfrak{g} \equiv \mathfrak{Lie}(G), \quad (14.123)$$

and this should also be the tensor branch of the 6d SCFT labeled by the corresponding Lie algebra $\mathfrak{g}$ *if it exists*. However to get *all* 5d SYM theories one does not need a distinct interacting 6d SCFT for each simple Lie algebra: the subset of *simply-laced* Lie algebras suffices. Indeed a non-simply-laced Lie algebra $\mathfrak{g}$ is the quotient of a simply-laced one $\mathfrak{g}_0$ with respect to some cyclic automorphism subgroup²⁷ generated by an element $\zeta \in \text{Aut}(\mathfrak{g}_0)$: this is the procedure we called “diagram folding” (cf. Fig. 2.2). Suppose that the 6d theory corresponding to $\mathfrak{g}_0$ exists. We can compactify it on a circle of angular coordinate θ with the twisted periodic condition²⁸

$$\mathcal{O}(\mathbf{x}, \theta + 2\pi) = \zeta \cdot \mathcal{O}(\mathbf{x}, \theta). \quad (14.124)$$

The effective 5d theory in the IR is maximal SYM with the non-simply-laced gauge group G . If we had an independent 6d SCFT associated to the non-simply-laced Lie algebra $\mathfrak{g}$, we would have *two* inequivalent UV completions of the 5d SYM with gauge group G : the compactification on S^1 of the 6d theory of type $\mathfrak{g}$ and the *twisted* compactification of the 6d theory of type $\mathfrak{g}_0$, a luxury which we ruled out as extremely unlikely. The most reasonable scenario is

²⁶ On the contrary for $d \leq 4$ we have typically several UV completions.

²⁷ The subgroup is $\mathbb{Z}_2$, except for the exceptional Lie algebra G_2 which is $D_4/\mathbb{Z}_3$.

²⁸ Here the symbol $\mathcal{O}(\mathbf{x}, \theta)$ stands for a general local operator of the 6d theory, and $\zeta \cdot$ for the action of the automorphism group on the local operators.

We have one (strongly) interacting 6d (2, 0) SCFT per each simply-laced simple Lie algebra A_r, D_r, E_6, E_7, E_8 ($r \geq 1$)

This conclusion is correct.²⁹ To prove the claim we have to show two points:

- (i) for each simply-laced $\mathfrak{g}$ we can construct a 6d (2, 0) SCFT;
- (ii) a part for the ones constructed in (i) there are *no other* 6d (2, 0) SCFT.

For (i) we shall sketch at the end of this section an explicit construction of the required SCFTs in string theory. Before going to that we discuss item (ii) and some other important physical points about the 6d (2, 0) SCFTs. Our discussion will be elementary (even a bit heuristic); readers looking for a deeper treatment may have a look to the quoted references.

Geometric SUSY Considerations

We stress that tensor multiplets are the only matter³⁰ (2, 0) light supermultiplet.

The tensor branch $\mathcal{T}$ is the scalars' target space of the low-energy effective theory in vacua where the strings are massive (hence integrated away). These vacua preserve 16 (Poincaré) supercharges. $\mathcal{T}$ parametrizes such SUSY vacua and is smooth except (possibly) at points where some string gets tensionless. As explained in Sect. 8.1, in absence of gravity, for more than 8 supercharges the scalars' manifold should be flat.³¹ Hence the metric completion of the universal cover of the smooth locus $\mathcal{T}_{\text{smooth}} \subset \mathcal{T}$ is $\mathbb{R}^{5r}$ with its standard flat metric³² (r being the number of tensor multiplets which are everywhere light on $\mathcal{T}$). Thus the metric completion of $\mathcal{T}$ has the form

$$\overline{\mathcal{T}} = \mathbb{R}^{5r} / \Gamma \quad (14.125)$$

where $\Gamma \subset \text{Iso}(\mathbb{R}^{5r})$ is a discrete group of isometries which may have fixed points.

Conic Geometries for SCFT However, our 6d theory is not just supersymmetric, is superconformal. In this case the scalars' manifold should be a *cone* over some Riemannian base B (see Sect. 7.5 of [48]), that is, we can find coordinates where

$$ds^2 = dr^2 + r^2 ds_B^2. \quad (14.126)$$

Indeed the symmetry generators include the dilatation D (i.e. the radial Hamiltonian; cf. Sect. 2.2.2) which acts on the scalars' kinetic metric G_{ab} as a CKV $V \equiv V^a \partial_a$ satisfying

$$\mathcal{L}_V G_{ab} = 2 G_{ab}. \quad (14.127)$$

²⁹ See, e.g. [78–93] for further elaboration.

³⁰ A *matter* supermultiplet is a representation of SUSY in terms of local fields with spin ≤ 1 .

³¹ For more details see Chap. 4 of [48].

³² This follows from Theorem 11.4.

However in $d \geq 3$ dimensions the existence of such a CKV is not enough: the free scalar Lagrangian $\partial^\mu \phi \partial_\mu \phi$ is not conformal since its energy-momentum tensor is not traceless: to get conformal invariance one should introduce an improvement term or, equivalently, a coupling to the scalar curvature R of spacetime.

Exercise 14.1 Show that the Gaussian conformal scalar theory in d dimensions has Lagrangian

$$\mathcal{L} = \frac{1}{2} \left(\partial^\mu \phi \partial_\mu \phi + \frac{(d-2)}{4(d-1)} R \phi^2 \right). \quad (14.128)$$

Hint. Use Box 1.12.

By the same token, the scalars of a general CFT in $d \geq 3$ should have a coupling to the spacetime scalar curvature R of the form $\frac{(d-2)}{2(d-1)} R \Phi$ for some globally-defined function Φ on the target space having the scaling property

$$\mathcal{L}_V \Phi = 2 \Phi \quad \text{and} \quad G_{ab} = D_a \partial_b \Phi \quad (14.129)$$

as one checks in the example (14.128). The gradient of Φ , $V_a = \partial_a \Phi$, is a CKV satisfying the equation

$$G_{ab} = D_a V_b \quad (14.130)$$

stronger than Eq. (14.127). Such a vector field is called a *concurrent*.³³

Lemma 14.1 *A Riemannian manifold is a cone (i.e. its metric has the form (14.126) in suitable local coordinates) iff it admits a concurrent CKV. In particular: the scalars' target space of all CFTs in $d \geq 3$ are Riemannian cones.*

Comparing with (14.125) we see that the tensor branch of a 6d (2, 0) SCFT must be a Riemannian cone over a base

$$S^{5r-1}/\Gamma \quad \text{with} \quad \Gamma \subset SO(5r) \text{ discrete, hence finite.} \quad (14.131)$$

In a SCFT the R-group $Sp(2)_R \simeq SO(5)$ should be an exact symmetry. This forces Γ to be a subgroup of the centralizer $SO(r)$ of $SO(5)$ in $SO(5r)$. Then Γ acts linearly on the tensor multiplet scalars and by SUSY this linear action extends to all tensor-multiplet fields including the 2-form gauge fields. The fluxes of the gauge field-strengths are quantized *à la* Dirac and only the arithmetic subgroup

$$SO(r; \mathbb{Z}) \subset SO(r) \subset SO(r) \times Sp(2)_R \simeq SO(r) \times SO(5) \subset SO(5r) \quad (14.132)$$

of the centralizer $SO(r) \subset SO(5r)$ is an actual symmetry, so that $\Gamma \subset SO(r; \mathbb{Z})$, i.e. Γ is a rational crystallographic subgroup in r dimensions. The tensor branch metric is singular at the tip of the cone, unless Γ is the trivial group (in which case the SCFT is free). However physics set restrictions on the possible singularities. The tensor-branch is parametrized by the v.e.v. of operators, and the ring of functions on $\mathcal{T}$,

³³ In the coordinates (14.126) one has $\Phi = r^2/2$.

$\mathbb{C}[x_1, \dots, x_{5r}]^\Gamma$ is isomorphic to a SUSY protected ring of local physical operators of the SCFT akin to the chiral rings studied in Chap. 11. It appears that physics requires this SUSY protected ring to be *regular* (in the sense of Commutative Algebra [94]). By the Sheppard–Todd–Chevalley theorem [95–101], the ring $\mathbb{C}[x_1, \dots, x_{5r}]^\Gamma$ is regular iff Γ is a finite reflection group. Since it is contained in $SO(r, \mathbb{Z})$, Γ is a crystallographic (finite) real reflection group of degree- r . It is well-known that they are precisely the Weyl groups of the rank- r (finite dimensional) Lie algebras [96–98]. We have recovered our guess (14.123). All our arguments, up to now, apply both to the 6d tensor branch and the 5d Coulomb branch (since we used only that we have 16 supersymmetries and the R-symmetry is $\mathrm{Sp}(2)_R$), and we did not get any restriction on the Lie algebra since there are none in 5d. What is specific to 6d is the *self-duality condition*, i.e. the identification of electric and magnetic charges and sources. This requires the Cartan matrix to be symmetric, i.e. the Lie algebra to be simply-laced. From this elementary geometric argument we learn that the 6d (2, 0) SCFT which may possibly exist are classified by the *ADE* Lie algebras, and no other one exists. Next we address the existence issue.

14.5.1 Construction of 6d (2, 0) SCFT from IIB on $\mathbb{C}^2/\Gamma$

IIB on a Local 2-CY In Sect. 13.3, we saw that the compactification of IIA on K3 is equivalent (dual) to the compactification of the heterotic string on T^4 . There we also discussed the non-Abelian enhancement of gauge symmetry at special points in the K3 moduli space that is the IIA dual description of the non-Abelian enhancement at special points in the moduli space of heterotic toroidal compactifications described in detail in Sect. 7.7. In the language of IIA on special K3's the light W -bosons are 2-branes wrapped on spheres of vanishing size, whose masses are given by the M2 tension τ_2 times their volume.

The discussion in Sect. 13.3 was local in a neighborhood of a crepant (hence canonical) singularity of K3, locally modeled on $\mathbb{C}^2/\Gamma$, with $\Gamma \subset SU(2)$ a finite subgroup (cf. Appendix 2 of Chap. 2). The gauge symmetry enhancement is conveniently studied in the local geometry $\mathbb{R}^{5,1} \times \mathbb{C}^2/\Gamma$, i.e. in IIA reduced to six-dimension using the non-compact ‘internal space’ $\mathbb{C}^2/\Gamma$ (dubbed a *local Calabi–Yau 2-fold*). Since the effective gravitational coupling κ_{eff}^2 in 6d is proportional to the inverse of the internal 4-volume, the non-compact space $\mathbb{C}^2/\Gamma$ yields $\kappa_{\mathrm{eff}}^2 = 0$, i.e. gravity is effectively decoupled, and we get a non-gravitational QFT in 6d.³⁴ The internal space $\mathbb{C}^2/\Gamma$ is HK of complex dimension 2, so according to Corollary 11.3 it has two parallel spinors of the same chirality, and the two Majorana–Weyl 10d supersymmetries of *opposite* chirality of IIA lead to two symplectic Majorana–Weyl 6d supercharges of opposite chirality, i.e. to (1, 1) SUSY in the language of Sect. 8.1 and Box 13.3.

The 6d IR QFT of IIA on $\mathbb{C}^2/\Gamma$ is then a (non-gravitational) 6d (1, 1) SUSY gauge theory with gauge group G the one associated to $\Gamma \subset SU(2)$ by the McKay

³⁴ This is akin to the fact, observed in Sect. 12.6.2, that the 9-9 strings d.o.f. decouple from the 6d world-volume QFT in a m D5- n D9 brane system.

correspondence. The W -bosons ($\equiv$ gauge vectors associated to the roots of the gauge algebra) arise from D2-branes wrapped on the $\mathbb{P}^1$'s of the crepant resolution $X_\Gamma \rightarrow \mathbb{C}^2/\Gamma$ of the surface singularity. We note that these $\mathbb{P}^1$'s are exceptional divisors in the resolved complex surface³⁵ X_Γ , hence the harmonic representatives of their cohomology classes $\eta_\alpha \in H^2(X_\Gamma, \mathbb{Z})$ are anti-self-dual, $\eta_\alpha = -*\eta_\alpha$, by Hirzebruch signature theorem [24].

IIB on a Local 2-CY Now we replace IIA with IIB. We still get a non-gravitational quantum theory in six-dimensions. However now the two 10d MW supersymmetries have the same chirality, hence the two 6d symplectic Majorana–Weyl SUSYs are of the *same* chirality, i.e. we have 6d (2, 0) *chiral* supersymmetry. The effective 6d QFT cannot be a gauge theory because, as discussed in Box 13.3, (2, 0) SUSY does not have light vector supermultiplets. Rather it has *tensor supermultiplets* whose sources are *strings* which become massless as we approach a singular point in the moduli space of the internal space. Moreover (2, 0) SUSY *predicts* that the 2-form in the tensor supermultiplet has a (anti)self-dual field-strength $H^{(3)}$.

To figure out the 6d SCFT which is the IR fixed point of IIB on $\mathbb{R}^{5,1} \times \mathbb{C}^2/\Gamma$, we start from the already understood case of IIA in the same geometry, and compactify it on a circle of radius R . In the IR this is a 5d $\mathcal{N} = 2$ SYM whose W -bosons are D2's wrapped on the exceptional $\mathbb{P}^1$'s in X_Γ . Now we exploit the fact that IIA and IIB compactified on a circle are equivalent under T -duality; cf. Sect. 12.1. T -duality along S_R^1 transforms a D2 wrapped on $\mathbb{P}^1 \subset X_\Gamma$ into a D3 wrapping

$$S_R^1 \times \mathbb{P}^1 \subset S_{R'}^1 \times X_\Gamma. \quad (14.133)$$

The D3 is a self-dual source for the self-dual 5-form field-strength $F^{(5)}$. We have

$$F^{(5)} \approx \sum_\alpha H_\alpha^{(3)} \wedge \eta_\alpha^{(2)} \quad (14.134)$$

where $H_\alpha^{(3)}$ are field-strength 3-forms in 6d and $\eta_\alpha^{(2)}$ are harmonic forms Poincaré-dual to the exceptional $\mathbb{P}^1$'s in the resolution X_Γ . Since both $F^{(5)}$ and $\eta_\alpha^{(2)}$ are (anti)self-dual, so are the $H_\alpha^{(3)}$ in agreement with the predictions from (2, 0) SUSY.

By T -duality the mass of the wrapped self-dual light string in IIB is equal to the mass of the corresponding W boson in IIA. Using the Buser rule for the transformation of the dilaton ($\equiv$ string coupling) under T -duality,³⁶ or the recursion relation for the D-brane tensions τ_p , Eq. (6.292), we see that

$$m_W = \tau_2 \text{vol}(\mathbb{P}^1) = 2\pi R' \tau_3 \text{vol}(\mathbb{P}^1) \quad (14.135)$$

as it should for a D3 wrapped on $S_{R'}^1 \times \mathbb{P}^1$.

³⁵ For this Hirzebruch signature argument we think of the local 2-CY X_Γ as an open subset of a global compact 2-CY.

³⁶ Cf. Sect. 6.4.2.

This construction in Type IIB shows that in 6d we have precisely one interacting (2, 0) SCFT per finite subgroup of $SU(2)$, i.e. per simply-laced ADE Lie algebra, as was to be proven. The 6d SCFTs have the expected properties, for instance reduce to SYM upon compactification.

Toroidal Compactifications. Montonen–Olive Duality Again

We have seen that the (untwisted) compactification of a 6d (2, 0) CFT on a circle reduces in the IR to the 5d SYM with the corresponding gauge Lie algebra. We wish to understand how the effective YM coupling in 5d, g_{YM}^2 , scales with the radius R of the circle.

We start with IIA on $\mathbb{C}^2/\Gamma$ times a circle of radius R_A : in 6d we have the standard SYM action

$$\frac{1}{4g_6^2} \int d^6x \left(\text{tr}(F^{(2)})^2 + \dots \right) = \frac{2\pi R_A}{4g_6^2} \int d^5x \left(\text{tr}(F^{(2)})^2 + \dots \right) \quad (14.136)$$

so that, in terms of the IIA quantities $g_{\text{YM}}^2 = g_6^2/2\pi R_A$, or in terms of the T -dual radius $R_B \equiv \alpha'/R_A$ of IIB

$$g_{\text{YM}}^2 \propto R_B \quad \Rightarrow \quad \mathcal{L}_{5d} \propto \frac{1}{R_B} \left(\text{tr} F_{\mu\nu} F^{\mu\nu} + \dots \right). \quad (14.137)$$

This equation implies that the 6d (2, 0) SCFT cannot possess a meaningful Lagrangian.³⁷ Indeed suppose *there was* a 6d Lagrangian $\mathcal{L}_{6d}$: one would get

$$\mathcal{L}_{5d} = \int_{S^1} dx \mathcal{L}_{6d} \sim 2\pi R_B \mathcal{L}_6 \Big|_{\text{zero mode}} \quad (14.138)$$

so the 5d YM coupling g_{YM}^2 would be proportional to R_B^{-1} instead of R_B : *the YM coupling scales the other way around with respect to Lagrangian field theory!* This, of course, reflects the fact that the fundamental light d.o.f. in 6d are now self-dual strings not point-particles as in Lagrangian field theory. Despite not having a Lagrangian description, the 6d theory is a genuine local QFT since it has a conserved *local* energy-momentum tensor.

Now let us compactify the 6d (2, 0) theory to 4d on a 2-torus $(S^1)^2$ of radii R_1 and R_2 , getting 4d $\mathcal{N} = 4$ SYM with the corresponding gauge group G . We can perform the compactification in two steps. In the first step, we get the effective 5d action

$$\propto \frac{1}{R_1} \int d^5x \left(\text{tr} F_{\mu\nu} F^{\mu\nu} + \dots \right) \quad (14.139)$$

and in the second step the 4d $\mathcal{N} = 4$ action

³⁷ By “meaningful” we mean a Lagrangian with the usual physical properties which makes it useful for computations and physically sound.

$$\propto \frac{R_2}{R_1} \int d^4x \left(\text{tr } F_{\mu\nu} F^{\mu\nu} + \dots \right). \quad (14.140)$$

We got a 4d YM coupling $g_{\text{YM}}^2 \propto R_1/R_2$. Of course we could have inverted the order of the compactifications, getting for the *same* physical system

$$g_{\text{YM}}^{2'} \propto R_2/R_1 \propto 1/g_{\text{YM}}^2. \quad (14.141)$$

These two values of the coupling are related by the strong-to-weak coupling S -duality of Montonen and Olive (cf. Sect. 13.1.2). Here we recover it as an obvious geometrical fact: the two S -dual couplings correspond to different descriptions of the *same* toroidal compactification, hence of the same quantum model. More generally: the mapping class group of the torus, $SL(2, \mathbb{Z})$ (cf. Sect. 5.1), acts on the cycles of T^2 which we may use to compactify by steps, so that the mere existence of the 6d $(2, 0)$ predicts that 4d $\mathcal{N} = 4$ SYM has a $SL(2, \mathbb{Z})$ group of non-perturbative dualities which in the 6d vantage perspective becomes a manifest geometric symmetry.

★ **4d $\mathcal{N} = 2$ SCFT of Class $\mathcal{S}$** More generally we can compactify a 6d $(2, 0)$ SCFT to 4d on a general Riemann surface Σ (with or without punctures) performing a partial topological twist to preserve some supersymmetry. In this way one constructs 4d $\mathcal{N} = 2$ QFTs, and in fact $\mathcal{N} = 2$ SCFT if the boundary conditions at the punctures are regular enough. Such 4d model are called $\mathcal{N} = 2$ SCFT of class $\mathcal{S}$ [102, 103]. “Most” class $\mathcal{S}$ theories do not admit a Lagrangian formulation. Again the mapping class group of the (punctured) surface Σ is physically realized as a group of non-perturbative dualities of the class- $\mathcal{S}$ $\mathcal{N} = 2$ SCFT [102].

14.6 Quantum Physics of Black Holes

Superstring theory is a *consistent* theory of Quantum Gravity. In particular, it should be able to escape the paradoxes one encounters in the field theoretic treatment of gravity at the quantum level. The most puzzling paradoxes arise from the Black Holes (BH) thermodynamics. In any number of dimensions a Black Hole has an entropy. For parametrically *large* Black Holes, the higher curvature interactions and quantum effects may be ignored and the semiclassical Bekenstein–Hawking formula [104–107] becomes reliable:

$$S \approx \frac{A}{4G_N} = \frac{2\pi A}{\kappa^2} \quad A \equiv \left[\begin{array}{l} \text{volume of horizon in} \\ \text{Einstein frame metric.} \end{array} \right. \quad (14.142)$$

This aspect of BH physics is puzzling since, in any quantum theory, unitarity requires $\exp(S)$ to be the number of quantum states and in particular a positive integer. To reconcile gravity and quantum physics we are forced to conclude that a large Black Hole of given energy (mass) and charges should be a microcanonical ensemble containing a huge number of quantum states (called *microstates*). This scenario leads to a couple of baffling questions:

- (a) *Where the microstates come from?*
 (b) *Can we see them and count their number?*

In a field theory approach to Quantum Gravity both questions look hopeless. But when the Black Hole is realized in superstring theory, the answer to question (b) becomes **Yes, we can** [108]. The informal answer to question (a) is then: *They come from the hidden “extra” (compact) dimensions.*

The counting of the Black Hole microstates in superstring theory is done in the literature for several classes of Black Holes, for BPS ones as well as for (some) *non*-BPS ones, in various dimensions and number of unbroken supersymmetries in the low-energy effective SUGRA, see [108–116] and references therein.

In some lucky cases, like four-dimensional $\frac{1}{2}$ -BPS Black-Holes in Type II compactified on T^6 , we have an *exact* formula (as a $E_7(\mathbb{Z})$ -invariant function of the 56 electric and magnetic charges of the BH) for $\exp(S)$ which yields a positive integer ($\equiv$ the precise number of microstates) for all values of the charges, even very small ones, i.e. also for small BH where the non-perturbative quantum effects dominate over the semiclassical contribution (14.142) (see, e.g. [109, 111, 113]). As expected, the exact microstate multiplicity $\exp(S)$ is a very subtle Number-Theoretic function of the 56 electric and magnetic charges of 32-SUSY supergravity.

Here we shall consider just the very simplest example of such microstate counting (and only its asymptotic behavior for large BH charges, where the semiclassical computation of S is reliable): this is the original set-up by Strominger and Vafa [108].

Basic Idea In superstring theory there are objects which—in different regions of their parameter space—have two dual interpretations: as a “soliton” of the low-energy effective field theory or as a bound state of branes. We have already seen several instances of this paradigm: the Yang–Mills instantons versus the $\#_{\text{ND}} = 4$ bound states of D-branes (Sect. 12.8), the (p, q) 5-branes (Sect. 8.8), systems of D6’s versus Taub-NUT geometries (Sect. 14.1.1), *etc.* When the relevant objects are BPS, we can control them over the full parameter space as we did in Sect. 13.1 to understand the Type IIB D1-string at strong coupling. In particular, the number of BPS states of the object (for given values of the conserved charges) is independent of the particular region in moduli space, so it may be computed in any convenient regime where the analysis simplifies.

To apply this strategy we need to construct a BPS configuration of branes having a dual realization as a “solitonic” solution of SUGRA which is a smooth Black Hole geometry with a regular horizon. Then we may compute the number of BPS states of the BH from the dual weakly-coupled brane picture and check if it reproduces (for asymptotically large BH) the Bekenstein–Hawking entropy (14.142) as given by the horizon volume of the BH geometry.

The Set-Up

Following [108], we consider Type IIB compactified to 5d on

$$\mathbb{R}_t \times \mathbb{R}_{x^1, \dots, x^4}^4 \times S_y^1 \times T_{w^1, \dots, w^4}^4 \quad (14.143)$$

where the subscripts in each factor space stand for the respective coordinates. The volume of T^4 (measured with the asymptotic metric at spatial infinity) is V_4 , while the length of S^1 is L . We take L to be finite but *large*.

On the geometry (14.143) we consider a number $Q_1 \gg 1$ of D1-branes wrapped on S^1 and a number of $Q_5 \gg 1$ of D5-branes wrapped on $S^1 \times T^4$. We also assume the system to have a momentum $2\pi n/L$ (with $1 \ll n \in \mathbb{N}$) along the circle S^1 . The set-up is specified by the three quantum numbers (conserved charges) Q_1 , Q_5 , and n .

The solution to the low-energy effective e.o.m. is very similar to the one for NS5 or D-branes presented in Sect. 8.8, except that now it is specified by three harmonic functions of the coordinates $x^1, \dots, x^4$ which are transverse to all branes, that is, the spatial coordinates of the non-compact 5d effective spacetime. The harmonic functions H_1 , H_5 , H_p are in one-to-one correspondence with the three non-trivial charges D1, D5, and p_5 carried by the solution.

Multi-Charge Solutions Before the Lorentz boost in the S^1 direction, the structure of a BPS multi-charge solution is as follows: the Einstein frame metric is given by the sum of the flat metrics of each factor space in (14.143) multiplied by the product of the warp factors $H_1^{s_1} H_5^{s_5}$ of the several branes where, as in the one-charge BPS solution (8.130),

$$s_p = \begin{cases} \frac{p+1}{8} - 1 & \text{Dp wrapped on the factor space} \\ \frac{p+1}{8} & \text{Dp unwrapped on the factor space,} \end{cases} \quad (14.144)$$

while $\exp(2\Phi)$ is given by the product

$$\prod_{p=1,5} H_p^{(3-p)/2} \quad (14.145)$$

of the several one-charge expressions. Boosting the solution in the y direction amounts to the replacement

$$-dt^2 + dy^2 \rightsquigarrow -dt^2 + dy^2 + H_n(x^i)(dt + dy)^2, \quad (14.146)$$

with $H_n(x^i)$ a certain function of the transverse directions x^i . The harmonic functions H_1 and H_5 are (respectively) as in the D1 and D5 brane solitons, while $H_n(x^i)$ should also have the same structure (in particular to be harmonic in $\mathbb{R}^4$) since momentum in the compact direction is U -dual to a RR charge (cf. Eqs. (13.42), (13.78)).

The BPS Solution

In our set-up (14.143) the *Einstein frame* metric and dilaton then read

$$ds_E^2 = H_1^{-3/4} H_5^{-1/4} [-dt^2 + dy^2 + H_n(dt + dy)^2] + H_1^{1/4} H_5^{3/4} dx^i dx^i + H_1^{1/4} H_5^{-1/4} dw^m dw^m \quad (14.147)$$

$$\text{dilaton: } e^{-2\Phi} = H_5/H_1, \quad (14.148)$$

where (here $r^2 \equiv x^i x^i$)

$$H_1 = 1 + \frac{r_1^2}{r^2} \quad r_1^2 = \frac{(2\pi)^4 g Q_1 \alpha'^3}{V_4} \quad (14.149)$$

$$H_5 = 1 + \frac{r_5^2}{r^2} \quad r_5^2 = g Q_5 \alpha' \quad (14.150)$$

$$H_n = \frac{r_n^2}{r^2} \quad r_n^2 = \frac{(2\pi)^6 g^2 n \alpha'^4}{L^2 V_4}. \quad (14.151)$$

The horizon is at $r = 0$. The horizon looks as S^3 in 5d, that is, $S^3 \times S^1 \times T^4$ from a 10d perspective. Near the horizon the *spatial part* of the metric reads

$$\begin{aligned} & \left(\frac{r_1}{r}\right)^{-3/2} \left(\frac{r_5}{r}\right)^{-1/2} \left(\frac{r_n}{r}\right)^2 dy^2 + \left(\frac{r_1}{r}\right)^{1/2} \left(\frac{r_5}{r}\right)^{-1/2} dw^2 + \\ & + \left(\frac{r_1}{r}\right)^{1/2} \left(\frac{r_5}{r}\right)^{3/2} (dr^2 + r^2 d\Omega^2) = \\ & = \frac{r_n^2}{r_1^{3/2} r_5^{1/2}} dy^2 + r_1^{1/2} r_5^{3/2} \frac{dr^2}{r^2} + r_1^{1/2} r_5^{3/2} d\Omega^2 + \frac{r_1^{1/2}}{r_5^{1/2}} dw^m dw^m \end{aligned} \quad (14.152)$$

with $d\Omega^2$ the round metric on the unit 3-sphere. The volume of the horizon $r = 0$ is

$$A = \text{vol}(S^3) \times \text{length}(S^1) \times \text{vol}(T^4) = \frac{r_n L}{r_1^{3/4} r_5^{1/4}} \times 2\pi^2 r_1^{3/4} r_5^{9/4} \times \frac{r_1 V_4}{r_5}, \quad (14.153)$$

that is, using Eqs. (14.149)–(14.151) and (12.54),

$$A = \frac{1}{2} (2\pi)^7 g^2 \alpha'^4 (Q_1 Q_5 n)^{1/2} \equiv \kappa^2 (Q_1 Q_5 n)^{1/2}, \quad (14.154)$$

so that the Bekenstein–Hawking entropy of the BPS solution is

$$S_{\text{BH}} = \frac{2\pi A}{\kappa^2} = 2\pi (Q_1 Q_5 n)^{1/2}. \quad (14.155)$$

Counting Microstates

Now we change perspective, look at our Black Hole set-up as a bound system of D-branes, and count the number of its BPS states at weak string coupling.

We have a $\#_{\text{ND}} = 4$ brane system, invariant under 8 SUSYs, whose effective theory was studied in Sect. 12.6.2: the D1 world-sheet effective theory is a 2d SUSY gauge theory with gauge group $U(Q_1) \times U(Q_5)$ given by light 1-1 and 5-5 strings (respectively) and a full hypermultiplet in the bi-fundamental representation $(\mathbf{Q}_1, \bar{\mathbf{Q}}_5)$ from the 1-5 and 5-1 strings. We write X_i and Y_i for the scalars in the 2d $U(Q_1)$ and $U(Q_5)$ gauge supermultiplets (seen as matrices in the respective fundamental representations) and χ for the bi-fundamental hypermultiplet scalars.

The scalar potential is³⁸

$$V = \frac{1}{(2\pi\alpha')^2} |X_i \chi - \chi Y_i|^2 + \frac{g_1^2}{4} \sum_{A=1}^3 \text{tr}(D_1^A D_1^A) + \frac{g_5^2}{4V_4} \sum_{a=1}^3 \text{tr}(D_5^a D_5^a). \quad (14.156)$$

We know from Chap. 12 that for a *bound state* of D1's and D5's $\chi \neq 0$, that is, we are on the Higgs branch of the $U(Q_1)$ gauge theory living on the D1 world-sheet. From the discussion in Sect. 12.8 we have the identification

$$\textbf{Higgs branch} \equiv \frac{\text{Moduli space of framed } SU(Q_5) \text{ instantons of topological charge } Q_1}{\text{instantons of topological charge } Q_1} \equiv \mathcal{M}_{Q_5, Q_1}, \quad (14.157)$$

so that,

$$\dim_{\mathbb{R}}(\textbf{Higgsbranch}) \equiv \dim_{\mathbb{R}} \mathcal{M}_{Q_5, Q_1} = 4 Q_1 Q_5. \quad (14.158)$$

We conclude that, at very low energies, the effective theory on the D1 world-sheet is a SUSY σ -model with target space the hyperKähler moduli space of instantons.³⁹ Since the target space is hyperKähler, the world-sheet theory on the D1 is a (4, 4) SCFT with Virasoro central charge 3/2 times the real dimension of the target space (cf. Sect. 11.2), that is,

$$c = \frac{3}{2} \dim_{\mathbb{R}} \mathcal{M}_{Q_5, Q_1} = 6 Q_1 Q_5. \quad (14.159)$$

The bounds states which preserve $\frac{1}{4} \cdot 32 = 8$ SUSY (that is, the vacua of the 8 supercharges $\#_{\text{ND}} = 4$ system; cf. Sect. 12.6.2) may be seen as $\frac{1}{2}$ -BPS states of the 16 supercharges D1 world-sheet theory. By S -duality (Sect. 13.1) the D1 world-sheet theory has the same physics as the F1 one. As we saw in Sect. 7.8, from the F1 world-sheet perspective the $\frac{1}{2}$ -BPS states have the property that their right-movers or left-movers are in a “ground” state. Let us focus on states whose right-movers are Ramond ground states, so they are annihilated by

$$\tilde{L}_0 - c/24 \equiv \frac{1}{2}(H - P). \quad (14.160)$$

Then

$$E \equiv L_0 - c/24 \equiv \frac{1}{2}(H + P) = P. \quad (14.161)$$

³⁸ Here D_1^A , D_5^a are the D -terms (in the sense of 4d $\mathcal{N} = 2$ SUSY!) of the SYM living on the D1 and D5, respectively; equivalently they are the three *hyperKähler* momentum maps of the $U(Q_1)$ and $U(Q_5)$ gauge symmetries (respectively) acting on the hypermultiplet scalars χ .

³⁹ Recall from the final paragraphs of Sect. 12.8 that the framed moduli space of instantons is a hyperKähler manifold.

We need to count such states as a function of P . This is done using the Cardy formula Eq. (2.174). Returning to our 5d Black Hole set-up we get for the number of states with momentum P along S^1

$$\begin{aligned} N(P) &= \exp \left[\sqrt{\frac{\pi c}{3}} L E + o(\sqrt{E}) \right] \approx \\ &\approx \exp \left[(2\pi Q_1 Q_5 P L)^{1/2} \right] \equiv \exp \left[2\pi (Q_1 Q_5 n)^{1/2} \right] \end{aligned} \quad (14.162)$$

since $P = 2\pi n/L$. The corresponding entropy

$$S(n) \equiv \log N(P) \approx 2\pi (Q_1 Q_5 n)^{1/2} \quad (14.163)$$

exactly matches the (asymptotic) expression for the Black Holes entropy (14.155)! The unitary puzzle with microstates has been brilliantly solved by string theory. Notice that the number of microstates depends only on the quantized charges, but not on the moduli or any other continuous parameters, as it should be for an integer valued function.⁴⁰

This result is quite a *triumph* in our understanding of Quantum Gravity! For the first time in history the entropy of the Black Holes was given a quantitative interpretation in the context of a *unitary* theory of Quantum Gravity, showing (in particular) that the Bekenstein–Hawking formula [104–107] is consistent with unitarity (contrary to previous claims) but (of course) this magic works only for very specific theories while for *generic* “matter coupled to gravity” models we get inescapable inconsistencies. This state of affairs led to the *swampland program*, i.e. the quest to characterize the set of physical models “smart enough” to be able to reconcile gravity with unitary quantum physics.

As a closing remark, we note that the extra coordinates (besides the visible ones) played a fundamental role in the reconciliation of these two fundamental physical principles. It is plausible that all consistent Quantum Gravities should have some “hidden” dimension (See [117]).

References

1. Gibbons, G.W., Hawking, S.W.: Gravitational multi-instantons. Phys. Lett. B **78**, 430–432 (1978)
2. Besse, A.: *Einstein Manifolds*. Springer, Berlin (1987)
3. Bogomolnyi, E.B.: The stability of classical solutions. Sov. J. Nucl. Phys. (Engl. Transl.) **24** 449 (1976)
4. Manton, N., Sutcliffe, P.: *Topological Solitons*. Cambridge University Press, Cambridge (2004)

⁴⁰ Heuristically this may be seen as an additional argument against free continuous parameters in quantum gravity.

5. Cecotti, S., Gaiotto, D., Vafa, C.: tt^* geometry in 3 and 4 dimensions. JHEP **05**, 055 (2014). [arXiv:1312.1008](#)
6. Hawking, S.W., Ellis, G.F.R.: The Large Scale Structure of Space-Time. Cambridge Monographs on Mathematical Physics, Cambridge University Press, Cambridge (1973)
7. Eguchi, T., Hanson, A.J.: Selfdual solutions to Euclidean gravity. Ann. Phys. **120**, 82–105 (1979)
8. Eguchi, T., Gilkey, P.B., Hanson, A.J.: Gravitation, gauge theories and differential geometry. Phys. Rev. **66**, 215–393 (1980)
9. Hitchin, N.J.: Polygons and gravitons. Math. Proc. Cambridge Philos. Soc. **83**, 465–476 (1969)
10. Kronheimer, P.B.: The construction of ALE spaces as hyper-Kähler quotients. J. Differ. Geom. **29**, 665–683 (1989)
11. Kronheimer, P.B.: Instantons gravitationnels et singularités de Klein. C. R. Acad. Sci. Paris Sér. I Math. **303**, 53–55 (1986)
12. Kronheimer, P.B.: A Torelli-type theorem for gravitational instantons. J. Differ. Geom. **29**, 685–697 (1989)
13. Reid, M.: The Du Val singularities A_n , D_n , E_6 , E_7 , E_8 . <https://homepages.warwick.ac.uk/~masda/surf/more/DuVal.pdf>
14. Reid, M.: La correspondance de McKay, Séminaire Bourbaki 1999/2000. Astérisque **276**, 53–72 (2002)
15. Milnor, J.: Singular Points of Complex Hypersurfaces. Annals of Math Studies, vol. 61. Princeton University Press, Princeton (1968)
16. Tachikawa, Y.: Frozen singularities in M and F theory. JHEP **06**, 128 (2016). [arXiv:1508.06679](#)
17. de Boer, J., Dijkgraaf, R., Hori, K., Keurentjes, A., Morgan, J., Morrison, D.R., Sethi, S.: Triples, fluxes, and strings. Adv. Theor. Math. Phys. **4**, 995–1186 (2002). [arXiv:hep-th/0103170](#)
18. Atiyah, M.F., Drinfeld, V.G., Hitchin, N.J., Manin, Y.I.: Phys. Lett. A **65**, 185 (1978)
19. Atiyah, M.F.: Geometry of Yang-Mills fields. Lezioni Fermiane, Scuola Normale Superiore, Pisa (1979)
20. Corrigan, E., Goddard, P.: Construction of instanton and monopole solutions and reciprocity. Ann. Phys. **154**, 253 (1984)
21. Tong, D.: TASI Lectures on Solitons: Instantons, Monopoles, Vortices and Kinks. [arXiv:hep-th/0509216](#)
22. Vandoren, S., van Nieuwenhuizen, P.: Lectures on Instantons. [arXiv:0802.1862](#)
23. Boyer, C., Galicki, K.: Sasakian Geometry. Oxford Mathematical Monographs, Oxford University Press, Oxford (2008)
24. Griffiths, P., Harris, J.: Principles of Algebraic Geometry. Wiley, Hoboken (1978)
25. Witten, E.: Sigma models and the ADHM construction of instantons. J. Geom. Phys. **15**, 215–226 (1995)
26. Douglas, M.R.: Gauge fields and D-branes. J. Geom. Phys. **28**, 255 (1998). [arXiv:hep-th/9604198](#)
27. Vafa, C.: Evidence for F-theory. Nucl. Phys. B **469**, 403–418 (1996). [arXiv:hep-th/9602022](#)
28. Morrison, D.R., Vafa, C.: Compactifications of F theory on Calabi-Yau threefolds. 1. Nucl. Phys. **B473** 74–92 (1996). [arXiv:hep-th/9602114](#)
29. Morrison, D.R., Vafa, C.: Compactifications of F theory on Calabi-Yau threefolds. 2. Nucl. Phys. **B476** 437–469 (1996). [arXiv:hep-th/9603161](#)
30. Weigand, T.: F-theory. PoS TASI2017, 016 (2018). [arXiv:1806.01854](#)
31. Cassels, J.W.S.: Lectures on Elliptic Curves. LMSST, vol. 24. Cambridge University Press, Cambridge (1991)
32. Lang, S., Néron, A.: Rational points of abelian varieties over function fields. Am. J. Math. **81**, 95–118 (1959)
33. Miranda, R.: The Basic Theory of Elliptic Surfaces. ETS Editrice, Pisa (1989)
34. Schütt, M., Shioda, T.: Mordell-Weil Lattices. Springer, Berlin (2019)
35. Aspinwall, P.S., Morrison, D.R.: Nonsimply connected gauge groups and rational points on elliptic curves. JHEP **07**, 012 (1998). [arXiv:hep-th/9805206](#)

36. Grimm, T.W., Weigand, T.: On Abelian gauge symmetries and proton decay in global F-theory GUTs. *Phys. Rev. D* **82**, 086009 (2010). [arXiv:1006.0226](#)
37. Kodaira, K.: On compact analytic surfaces. II. *Ann. Math.* **77**, 563–626 (1963)
38. Kodaira, K.: On compact analytic surfaces. III. *Ann. Math.* **78**, 1–40 (1963)
39. Barth, W., Hulek, K., Peter, C., van de Ven, A.: *Compact Complex Surfaces*, 2nd edn. *Ergebnisse der Mathematik und ihrer Grenzgebiete, 3. Folge, Band 4*. Springer, Berlin (2004)
40. Braun, V., Morrison, D.R.: F-theory on genus-one fibrations. *JHEP* **08**, 132 (2014). [arXiv:1401.7844](#)
41. Donagi, R., Wijnholt, M.: Model building with F-theory. *Adv. Theor. Math. Phys.* **15**, 1237–1317 (2011). [arXiv:0802.2969](#)
42. Beasley, C., Heckman, J.J., Vafa, C.: GUTs and exceptional branes in F-theory - I. *JHEP* **01**, 058 (2009). [arXiv:0802.3391](#)
43. Beasley, C., Heckman, J.J., Vafa, C.: GUTs and exceptional branes in F-theory - II: Experimental predictions. *JHEP* **01**, 059 (2009). [arXiv:0806.0102](#)
44. Donagi, R., Wijnholt, M.: Breaking GUT groups in F-theory. *Adv. Theor. Math. Phys.* **15**, 1523–1603 (2011). [arXiv:0808.2223](#)
45. Heckman, J.J.: Particle physics implications of F-theory. *Ann. Rev. Nucl. Part. Sci.* **60**, 237–265 (2010). [arXiv:1001.0577](#)
46. Kodaira, K.: On the structure of compact complex analytic surfaces. I. *Am. J. Math.* **86**, 751–798 (1964)
47. Schmid, W.: Variation of hodge structure: the singularities of the period mapping. *Inv. Math.* **22**, 211–319 (1973)
48. Cecotti, S.: *Supersymmetric Field Theories. Geometric Structures and Dualities*, Cambridge University Press, Cambridge (2015)
49. Bhardwaj, L., Morrison, D.R., Tachikawa, Y., Tomasiello, A.: The frozen phase of F-theory. *JHEP* **08**, 138 (2018). [arXiv:1805.09070](#)
50. Witten, E.: Toroidal compactification without vector structure. *JHEP* **02**, 006 (1998). [arXiv:hep-th/9712028](#)
51. Tate, J.: Algorithm for determining the type of a singular fiber in an elliptic pencil. *Lecture Notes in Mathematics*, vol. 476. Springer, Berlin (1975)
52. Bershadsky, M., Intriligator, K.A., Kachru, S., Morrison, D.R., Sadov, V., Vafa, C.: Geometric singularities and enhanced gauge symmetries. *Nucl. Phys. B* **481**, 215–252 (1996). [arXiv:hep-th/9605200](#)
53. Katz, S., Morrison, D.R., Schafer-Nameki, S., Sully, J.: Tate’s algorithm and F-theory. *JHEP* **08**, 094 (2011). [arXiv:1106.3854](#)
54. Anderson, L.B., Gao, X., Gray, J., Lee, S.J.: Fibrations in CICY threefolds. *JHEP* **10**, 077 (2017). [arXiv:1708.07907](#)
55. Denef, F.: Les Houches lectures on constructing string vacua. *Les Houches* **87**, 483–610 (2008). [arXiv:0803.1194](#)
56. Banks, T., Fischler, W., Shenker, S.H., Susskind, L.: M theory as a matrix model: a conjecture. *Phys. Rev. D* **55**, 5112–5128 (1997). [arXiv:hep-th/9610043](#)
57. Bigatti, D., Susskind, L.: Review of matrix theory. *NATO Sci. Ser. C* **520**, 277–318 (1999). [arXiv:hep-th/9712072](#)
58. Banks, T.: Matrix theory. *Nucl. Phys. B Proc. Suppl.* **67**, 180–224 (1998). [arXiv:hep-th/9710231](#)
59. Banks, T.: TASI Lectures on Matrix Theory. [arXiv:hep-th/9911068](#)
60. Seiberg, N.: Why is the matrix model correct? *Phys. Rev. Lett.* **79**, 3577–3580 (1997). [arXiv:hep-th/9710009](#)
61. Sen, A.: D0-branes on $T^{**}n$ and matrix theory. *Adv. Theor. Math. Phys.* **2**, 51–59 (1998). [arXiv:hep-th/9709220](#)
62. Banks, T., Seiberg, N., Shenker, S.H.: Branes from matrices. *Nucl. Phys. B* **490**, 91–106 (1997). [arXiv:hep-th/9612157](#)
63. Berkooz, M., Douglas, M.R.: Five-branes in M(atrix) theory. *Phys. Lett. B* **395**, 196–202 (1997). [arXiv:hep-th/9610236](#)

64. de Wit, B., Hoppe, J., Nicolai, H.: On the quantum mechanics of supermembranes. Nucl. Phys. B **305**, 545–81 (1988)
65. Cannas da Silva, A.: Lectures on Symplectic Geometry. Lectures Notes in Mathematics, vol. 1764. Springer, Berlin (2008)
66. Arnol'd, V.I.: The cohomology ring of the colored braid group. Math. Notes **5**, 138–140 (1969)
67. Connes, A.: Noncommutative Geometry. Academic, Cambridge (1995)
68. Kassel, C.: *Quantum Groups*. Graduate Texts in Mathematics, vol. 155. Springer, Berlin (1995)
69. Assem, I., Simson, D., Skowroński, A.: Elements of the Representation Theory of Associative Algebras. 1: Techniques of Representation Theory. London Mathematical Society Student Text, vol. 65. Cambridge University Press, Cambridge (2006)
70. Bourbaki, N.: Éléments de mathématique. Groupes et algèbres de Lie. Chapitre 1: Algèbres de Lie. Springer, Berlin (2007)
71. Groenewold, H.J.: On the principles of elementary quantum mechanics. Physica **12**, 405–460 (1946)
72. Curtright, T.L., Zachos, C.K.: Quantum mechanics in phase space. Asia Pac. Phys. Newslett. **1**, 37–46 (2012). [arXiv:1104.5269](https://arxiv.org/abs/1104.5269) [physics.hist-ph]
73. Zachos, C., Fairle, D., Curtright, T.: Quantum Mechanics in Phase Space. World Scientific, Singapore (2005)
74. Curtright, T., Fairle, D., Zachos, C.: A Concise Treatise on Quantum Mechanics in Phase Space. World Scientific, Berlin (2014)
75. Kontsevich, M.: Deformation quantization of Poisson manifolds. 1. Lett. Math. Phys. **66**, 157–216 (2003). [arXiv:q-alg/9709040](https://arxiv.org/abs/q-alg/9709040)
76. Landau, L.D., Lifshitz, E.M.: Quantum Mechanics. Non-relativistic Theory, 3rd edn. Elsevier, Amsterdam (1981)
77. Cecotti, S., Neitzke, A., Vafa, C.: R-twisting and 4d/2d correspondences. [arXiv:1006.3435](https://arxiv.org/abs/1006.3435)
78. Seiberg, N.: Nontrivial fixed points of the renormalization group in six-dimensions. Phys. Lett. B **390**, 169–171 (1997). [arXiv:hep-th/9609161](https://arxiv.org/abs/hep-th/9609161)
79. Witten, E.: Some comments on string dynamics. In: STRINGS 95: Future Perspectives in String Theory, vol. 7, pp. 501–523 (1995). [arXiv:hep-th/9507121](https://arxiv.org/abs/hep-th/9507121)
80. Strominger, A.: Open p -branes. Phys. Lett. B **383**, 44–47 (1996). [arXiv:hep-th/9512059](https://arxiv.org/abs/hep-th/9512059)
81. Gaiotto, O.J., Hanany, A.: Small $E(8)$ instantons and tensionless noncritical strings. Nucl. Phys. B **474**, 122–140 (1996). [arXiv:hep-th/9602120](https://arxiv.org/abs/hep-th/9602120)
82. Seiberg, N., Witten, E.: Comments on string dynamics in six-dimensions. Nucl. Phys. B **471**, 121–134 (1996). [arXiv:hep-th/9603003](https://arxiv.org/abs/hep-th/9603003)
83. Witten, E.: Phase transitions in M theory and F theory. Nucl. Phys. B **471**, 195–216 (1996). [arXiv:hep-th/9603150](https://arxiv.org/abs/hep-th/9603150)
84. Morrison, D.R., Vafa, C.: Compactifications of F theory on Calabi-Yau threefolds. 2. Nucl. Phys. B **476**, 437–469 (1996). [arXiv:hep-th/9603161](https://arxiv.org/abs/hep-th/9603161)
85. Witten, E.: Physical interpretation of certain strong coupling singularities. Mod. Phys. Lett. A **11**, 2649–2654 (1996). [arXiv:hep-th/9609159](https://arxiv.org/abs/hep-th/9609159)
86. Bershadsky, M., Johansen, A.: Colliding singularities in F theory and phase transitions. Nucl. Phys. B **489**, 122–138 (1997). [arXiv:hep-th/9610111](https://arxiv.org/abs/hep-th/9610111)
87. Brunner, I., Karch, A.: Branes at orbifolds versus Hanany Witten in six-dimensions. JHEP **03**, 003 (1998). [arXiv:hep-th/9712143](https://arxiv.org/abs/hep-th/9712143)
88. Blum, J.D., Intriligator, K.A.: New phases of string theory and 6-D RG fixed points via branes at orbifold singularities. Nucl. Phys. B **506**, 199–222 (1997). [arXiv:hep-th/9705044](https://arxiv.org/abs/hep-th/9705044)
89. Blum, J.D., Intriligator, K.A.: Consistency conditions for branes at orbifold singularities. Nucl. Phys. B **506**, 223–235 (1997). [arXiv:hep-th/9705030](https://arxiv.org/abs/hep-th/9705030)
90. Aspinwall, P.S., Morrison, D.R.: Point - like instantons on K3 orbifolds. Nucl. Phys. B **503**, 533–564 (1997). [arXiv:hep-th/9705104](https://arxiv.org/abs/hep-th/9705104)
91. Intriligator, K.A.: New string theories in six-dimensions via branes at orbifold singularities. Adv. Theor. Math. Phys. **1**, 271–282 (1998). [arXiv:hep-th/9708117](https://arxiv.org/abs/hep-th/9708117)

92. Hanany, A., Zaffaroni, A.: Branes and six-dimensional supersymmetric theories. *Nucl. Phys. B* **529**, 180–206 (1998). [arXiv:hep-th/9712145](#)
93. Heckman, J.J., Morrison, D.R., Vafa, C.: On the classification of 6D SCFTs and generalized ADE orbifolds. *JHEP* **05**, 028 (2014). [arXiv:1312.5746](#). [Erratum: *JHEP* **06**, 017 (2015)]
94. Eisenbud, D.: *Commutative Algebra with a View Toward Algebraic Geometry*. Graduate Texts in Mathematics, vol. 150. Springer, Berlin (1995)
95. Shephard, G.C., Todd, J.A.: Finite unitary reflection groups. *Can. J. Math.* **6**, 274–304 (1954)
96. Dolgachev, I.G.: *Reflection groups in algebraic geometry*. Bull. AMS (2008)
97. Gec, M., Malle, G.: Contribution to the handbook of algebra: reflection groups. [arXiv:math/0311012](#)
98. Lehrer, G.I., Taylor, D.E.: *Unitary Reflection Groups*. Cambridge University Press, Cambridge (2009)
99. Broué, M., Malle, G., Roquier, R.: Complex reflection groups, braid groups, Hecke algebras. *J. Reine Angew. Math.* **500**, 127–190 (1998)
100. Broué, M.: Introduction to complex reflection groups and their braid groups. *Lecture Notes in Mathematics*, vol. 1988. Springer, Berlin (2010)
101. Broué, M.: Reflection groups, braid groups, Hecke algebra, finite reductive groups. *Current Developments in Mathematics*, 2000, pp. 1–107. International Press, Vienna (2001)
102. Gaiotto, D.: $N = 2$ dualities. *JHEP* **08**, 034 (2012). [arXiv:0904.2715](#)
103. Gaiotto, D., Moore, G.W., Neitzke, A.: Wall-crossing, Hitchin systems, and the WKB approximation. *Adv. Math.* **234**, 239–403 (2013). [arXiv:0907.3987](#)
104. Bekenstein, J.D.: Black holes and the second law. *Lett. Nuovo Cim.* **4**, 737–740 (1972)
105. Bekenstein, J.D.: Black holes and entropy. *Phys. Rev. D* **7**, 2333–2346 (1973)
106. Bekenstein, J.D.: Generalized second law of thermodynamics in black hole physics. *Phys. Rev. D* **9**, 3292–3300 (1974)
107. Hawking, S.W.: Gravitational radiation from colliding black holes. *Phys. Rev. Lett.* **26**, 1344–1346 (1971)
108. Strominger, A., Vafa, C.: Microscopic origin of the Bekenstein-Hawking entropy. *Phys. Lett. B* **379**, 99–104 (1996). [arXiv:hep-th/9601029](#)
109. Maldacena, J.M., Moore, G.W., Strominger, A.: Counting BPS black holes in toroidal type II string theory. [arXiv:hep-th/9903163](#)
110. Ooguri, H., Strominger, A., Vafa, C.: Black hole attractors and the topological string. *Phys. Rev. D* **70**, 106007 (2004). [arXiv:hep-th/0405146](#)
111. Shih, D., Strominger, A., Yin, X.: Counting dyons in $N=8$ string theory. *JHEP* **06**, 037 (2006). [arXiv:hep-th/0506151](#)
112. Dabholkar, A., Sen, A., Trivedi, S.P.: Black hole microstates and attractor without supersymmetry. *JHEP* **01**, 096 (2007). [arXiv:hep-th/0611143](#)
113. Sen, A.: U -duality invariant Dyon spectrum in type II on T^{*6} . *JHEP* **08**, 037 (2008). [arXiv:0804.0651](#)
114. Sen, A.: Black Hole entropy function, attractors and precision counting of microstates. *Gen. Rel. Grav.* **40**, 2249–2431 (2008). [arXiv:0708.1270](#)
115. Mandal, I., Sen, A.: Black Hole microstate counting and its macroscopic counterpart. *Class. Quant. Grav.* **27**, 214003 (2010). [arXiv:1008.3801](#)
116. Sen, A.: Microscopic and macroscopic entropy of extremal Black Holes in string theory. *Gen. Rel. Grav.* **46**, 1711 (2014). [arXiv:1402.0109](#)
117. Montero, M., Vafa, C., Valenzuela, I.: The dark dimension and the swampland. *JHEP* **02**, 022 (2023). [arXiv:2205.12293](#)

Index

A

Abelian variety, 38, 589, 630, 631, 708
Abstruse identity, 331, 338, 342, 343, 348
ADHM construction, 714, 716, 779, 781–784
Ad-invariant polynomial, 528, 530, 537, 538
ALE space, 775
Anomaly polynomial, 527–533, 535–538, 540–544, 547, 569, 619, 620
Artin group, 313, 314
Artin–Tate group, 314
Atiyah–Singer index, 36, 531, 533

B

Barnes function, 455
Bekenstein–Hawking entropy, 813, 816
Beltrami differential, 41, 44, 559, 594
Bernoulli numbers, 455
Bernoulli polynomial, 20
 β -function, 53, 216, 284, 285, 476
 B -field, 22
Black branes, 509, 520, 727
Black hole, 6, 58, 394, 509, 520, 735, 813–818
Bootstrap, 80, 107
Borel density, 393, 462
Born–Infeld action, 415, 666, 667, 669, 671, 690, 692, 798
Bose sea, 116, 117
Bott periodicity, 169, 484
Boundary, 4, 5, 7, 10, 12, 24, 27, 34, 35, 52, 129, 130, 251, 252, 270, 304, 310, 326
Boundary conditions, 12, 16, 19, 20, 46, 77, 131, 152, 156, 159, 269, 275, 276,

281, 315, 407, 409, 414, 437, 657, 674, 693, 696

Boundary normal order, 270
Boundary states, 276, 304, 330, 419
BPS bound, 155, 474, 489, 510, 517, 658, 663, 689, 691–694, 701, 705, 797
BPS state, 473, 474, 488, 658, 661, 672, 675, 691, 694, 703, 706, 707, 711, 736, 737, 750, 814
Braid group, 313
Breathing mode, 628
Brinkmann space, 521, 522
Busher rules, 380, 381, 762, 796, 811

C

Canonical line bundle, 38, 47, 114, 570, 584, 587, 590, 592, 595, 777, 792
Canonical quantization, 14, 17, 18, 382, 466
Cardy formula, 98, 104, 818
Cartan algebra, 97, 164, 410, 807
Cartan–Hadamard theorem, 392, 467, 548
Cartan matrix, 134, 165, 185–188, 454, 778, 791, 793, 810
Cartan structure equations, 15
Cartan theorem A, 74
Casimir energy, 19, 20, 95, 159
Casimir invariant, 138, 139, 141, 145, 395, 457
Cayley transform, 129, 269
Central charge, 90–92, 104, 115, 120, 123, 135, 138, 145, 162, 174, 431, 473, 475, 491, 632
Chan–Paton factor, 252, 254, 290, 405, 552
Chan–Paton label, 7, 251, 252, 254, 255, 289, 291, 326, 341, 404, 406, 408, 412, 413, 422, 660, 710, 745, 746

Chern class, 37, 182, 353, 513, 515, 587, 588, 752, 760
 Chern–Simons form, 507, 530, 540, 555
 Chern–Simons term, 495, 498, 536, 539, 554, 563, 668, 704, 711, 744, 752
 Chiral algebra, 100, 101, 137, 173, 174, 177
 Chiral bosons, 324, 366, 501, 531, 533
 Chiral current, 81, 83, 100, 111, 112, 135, 136, 148, 173, 175, 459, 601, 612
 Chiral partition functions, 319
 Chiral scalar, 112, 120, 121, 124, 133, 147, 171, 366, 448, 449
 Class $\mathcal{S}$ SCFT, 813
 c -map, 644
 Cocycle, 37, 166, 168, 199, 366
 Coleman–Weinberg formula, 298, 335
 Complete (vector field), 68
 Complex automorphism, 28
 Complex structure, 24–26, 29, 40–42, 60, 102, 103, 182, 296, 313, 579, 590–594, 598, 610, 616, 772, 773, 788
 Concurrent vector, 809
 Cone, 808
 Conformal family, 98
 Conformal Killing spinor, 47, 148, 161
 Conformal Killing vector, 24, 27, 46, 72, 74
 Conformal motion, 24
 Conformal spin, 87
 Conformal structure, 40
 Connection form, 353, 355, 760, 771
 Contact term, 70, 81, 82, 92, 566
 Coulomb branch, 668, 710, 711, 713, 806, 807
 Coulomb gauge, 14, 206
 Coxeter element, 139
 Coxeter group, 314
 Coxeter label, 142, 143, 186, 440, 793
 Coxeter number, 138, 139, 440, 508
 Crepant resolution, 189, 708, 775, 778, 810, 811
 Cross-cap, 26, 132, 305, 306, 661
 Cross-cap state, 132, 304
 Curvature operators, 392

D
 D-brane, 410–417, 420
 D-brane tension, 414
 Decompactification limit, 375, 412, 468, 747, 798
 Dedekind function, 273, 274, 331, 336
 Deformation quantization, 803
 Dehn twist, 314–317, 319

Descendent fields, 87
 Diagonal invariant, 340
 Diffeomorphism group, 11
 Diffeomorphisms, 11, 25, 27, 68, 93, 309, 310, 352, 591, 593, 789, 801
 Dilaton, 22, 51, 52, 54, 55, 122, 295, 326, 378, 381, 499
 Dilaton tadpole, 326, 343, 413
 Dirac–Ramond operator, 213, 218, 543
 Dirichlet b.c., 13, 77, 79, 304, 407, 409, 411, 417, 418, 657, 662, 664, 672, 736
 Dirichlet boundary state, 304
 Dirichlet brane, 411
 Dirichlet integral, 51
 Discrete torsion, 402, 446, 447
 Discriminant, 787
 Discriminant function, 275
 Dolbeault differential, operator, 585
 Dolbeault–Kähler–Dirac operator, 584
 Double cover, 26
 Double-gamma function, 455
 Doubling trick, 131, 132, 152, 218, 226, 269, 275, 280–282, 330, 334, 674, 739
 Dualizing bundle/sheaf, 38

E

E_8 , 454
 E_8 matrix, 456
 Eguchi–Hanson instanton, 775
 Einstein frame, 56, 392, 419, 496, 502, 503, 505, 507, 512, 517, 518, 724, 730, 796
 Elliptic Calabi–Yau, 794
 Elliptic genus, 542
 End of the world wall, 763
 Euler beta-function, 284, 285
 Euler characteristics, 9, 10, 27, 38, 590, 708, 794

F

Faddeev–Popov determinant, 28
 Faddeev–Popov ghosts, 30
 Fermi sea, 116, 117
 Field strength, 52
 Flow (of a vector field), 68
 Flux-tube, 3
 Frenkel–Kač–Segal construction, 146, 162, 164, 166, 440
 Frobenius–Schur indicator, 746
 F-theory, 785–788, 790–792, 794
 Fundamental domain, 102, 277, 296, 297

G

Gauss–Bonnet theorem, 10, 43, 122
 Geodesic curvature, 10
 Ghost number, 36
 Gibbons–Hawking metric, 770–776
 Global gravitational anomalies, 309
 Goldstino, 657
 Grand anomaly polynomial, 543
 Graviton, 471, 623
 Gravitational instanton, 770
 Gravity, 6, 22
 Green–Schwarz mechanism, 527, 538, 539, 541
 Ground state, 21
 GSO projection, 157, 163, 170, 171, 200, 204, 205, 212, 222, 243, 250, 322, 323, 336, 341, 370, 401, 432–434, 440, 459

H

Hadamard space, 392, 461, 464, 467
 Hardy–Ramanujan formula, 98
 Harish–Chandra isomorphism, 530
 Heterotic string, 429–435, 437–440, 443, 445–448
 Higgs branch, 710, 712, 713, 716, 744, 779, 782, 784, 785, 817
 Highest root, 135, 139, 142
 Hodge decomposition, 365, 586, 596, 631, 636–638
 Hodge diamond, 586
 Hodge norm, 361, 364, 504, 505, 724
 Holonomy, 25, 404, 470, 476, 485, 501, 514, 521, 527, 530, 579–582, 584, 587, 588, 592, 623, 627, 656, 679, 712, 759, 770
 Holonomy group, 580
 Holonomy principle, 581, 582
 Hopf surface, 586
 Hurwitz ζ -function, 20
 Hyperelliptic curve, 398
 HyperHamiltonian action, 715
 Hyperholomorphic bundle, 773
 HyperKähler quotient, 715, 775

I

Improved field strength, 357, 358, 495, 496, 502
 Index theorem, 36, 527, 531, 532, 544
 Infinitesimal braid relations, 145
 Inheritance principle, 398
 Instanton, 36, 653, 660, 710–714

Integral basis, 452

J

Jacobi triple product, 278, 329, 331, 368, 545, 684

K

Kähler, 24, 25, 28, 33, 43, 476, 477, 486, 579–582, 584–587, 592, 596, 599
 Kähler cone, 592
 Kähler–Dirac, 500, 513, 535
 Kaluza–Klein geometry, 351–353, 355, 769
 Kaluza–Klein mechanism, 352, 354–356, 776
 Khono connection, 144
 Killing vector, 68
 KK monopole, 353, 760
 Klein bottle, 27, 276, 330
 Knizhnik–Zamolodchikov equation, 144
 Kodaira theorem, 26
 Kosterlitz–Thouless transition, 367, 403
 K3 surface, 23, 589, 590, 737, 738, 778, 794

L

Lattice, 449, 453
 Left-movers, 16
 Left-right matching, 21
 Level, 21, 135
 Level operator, 208
 Lie derivative, 69
 Light cone, 14
 Linear dilaton, 55, 122, 123
 Liouville QFT, 33
 Liouville superfield, 45
 Local CY, 811
 Loop algebras, 133
 Lorentz symmetry, 21

M

Magnetic monopole, 353
 Majorana–Weyl, 45
 Mandelstam variables, 284, 286
 Mapping class group, 310, 389
 Marginal operators, 367
 Marking (of a surface), 311
 Marsden–Weinstein quotient, 715
 Massive IIA SUGRA, 499, 500
 Matrix theory, 797, 798
 Maurier–Cartan form, 299, 390, 502
 M2 brane, 750
 M5 brane, 750

McKay correspondence, 188, 775, 777
 Meson, 3
 Microstates, 813
 Milnor theorem, 190, 778
 Minimal models, 100, 101
 Minimal models ($\mathcal{N} = 2$), 612
 Minkowski–Siegel formula, 454
 Möbius strip, 278
 Möbius transformation, 94, 313
 Modular group, 102, 275, 311, 312, 319, 364, 520, 725, 726, 749
 Modular invariance, 102–104, 169, 312, 315, 317, 319, 321, 328, 335, 337, 338, 340, 361, 385, 387, 401, 435, 450, 451, 542, 543, 546, 547
 Moduli space, 29, 40–42, 44, 60, 62, 102, 216, 265, 279, 295, 301, 313, 335, 377, 388, 391, 392, 395, 461, 464–467, 472, 548, 556, 590, 592, 594, 596, 599, 617, 618, 630, 631, 641, 642, 645, 712–714, 735, 738, 739, 765, 779, 787, 817
 Modulus (of torus), 295
 Momentum, 16
 Montonen–Olive duality, 728, 812
 Mordell–Weil group, 787
 Moyal product, 803
 M-theory, 623, 739, 746, 748–750, 753–760, 763–766, 776–778, 789
 Multi-center Taub–NUT, 775
 Multi-charge solutions, 815
 Multi-critical points, 367
 Mutual locality, 156

N
 Nambu–Goto action, 8, 414
 Narain moduli space, 389, 391, 392, 395, 461, 464–468
 Neumann b.c., 12, 269, 275, 276, 278, 334, 407, 409, 413, 418, 662, 672, 676, 736
 Neumann boundary state, 304
 Neveu–Schwarz/Ramond superstring, 44
 Neveu–Schwarz sector, 151
 No-ghost theorem, 211, 224, 232, 251, 255
 Non-oriented, 4, 7, 16, 27, 129, 132, 253–255, 265, 307, 423
 Non-renormalization theorem, 488, 569
 Null state, 99, 209

O
 Old covariant quantization, 206

Operator product expansion, 79, 80, 90–92, 97, 107, 110–112, 115, 119, 136, 138, 140, 141, 148–150, 152, 156–158, 170, 174, 177, 178, 198, 280, 339, 367, 374, 441, 602, 611
 Orbifold, 396
 Orbit space, 352
 Order of transitivity, 67
 Oriented, 4
 Oriented double, 24–27, 276, 282, 304
 Orientifold, 421
 Oscillator number, 19
 Oscillator operators, 18

P

Parallel forms, 581
 Parallel spinors, 582
 Partitions, 98
 Period domain, 632
 Periods, 311
 Picard group, 37, 85, 591
 Picture changing, 556, 559
 Picture charge, 125, 126, 200, 220, 229, 234, 235, 613
 Pluri-subharmonic function, 592
 Pochhammer’s symbol, 286
 Poincaré–Lelong formula, 92
 Poincaré symmetry, 10
 Poisson summation formula, 361, 363, 384, 386
 Polarization, 630
 Polyakov action, 9, 10
pp-waves, 521
 Primary fields, 86
 Projective curve, 26
 Projective plane (real), 26
 Punctures, 40
 Pure mapping class group, 310

Q

Quadratic differentials, 41
 Quantum torus algebra, 802
 Quark, 3
 Quasi-topological d.o.f., 232, 245, 246, 344, 499, 500

R

Radial quantization, 76
 Ramond sector, 151
 Rarita–Schwinger, 490
 Rational Conformal Field Theory, 101

Reduced space, 557
 Refermionization, 172
 Reflection group, 810
 Regge behavior, 288, 292
 Regge limit, 287
 Regge slope, 8, 288, 670
 Reissner-Nordström BH, 509
 Ricci tensor, 33, 57, 299, 477, 511, 582, 583, 587, 597, 622, 625, 627, 628
 Riemann-Hilbert correspondence, 404
 Riemannian cone, 808
 Riemann relations, 368
 Riemann-Roch theorem, 36, 39, 122, 561, 570
 Riemann surface, 24
 Riemann ζ -function, 455
 Right-movers, 16
 Romans mass, 500

S

Satake boundary, 725
 Scalar curvature, 10
 Scaling dimension, 87
 Scaling operator, 76
 (2, 0) SCFT, 806
 Schottky locus, 311
 Schrödinger representation, 78
 Schwarz reflection principle, 131
 Serre duality, 38, 570
 Shift, 19
 Short supermultiplet, 490
 Siegel domain, 599
 Siegel modular group, 311
 Siegel upper half-space, 311
 σ -model, 51
 Singular state, 99
 $SL(2, \mathbb{Z})$ duality, 725
 Spacetime, 4
 Special geometry, 629
 Spectral flow, 328
 Spectrum-generating algebra, 260
 Spin & statistics, 340
 Spin fields, 156, 157, 167, 170–172, 326, 340, 654
 Spin structure, 46–48, 543, 570
 Spin-2, 5
 Spinorial weights, 167
 Split supermanifold, 557
 Spurious states, 209
 State-operator correspondence, 76, 78, 84, 87, 96, 98, 100, 121, 128, 137, 154, 179, 200, 214, 217, 221, 235, 238, 248, 294, 316, 320, 344, 610

Stein manifold, 74
 String frame, 56
 String length, 8
 String tension, 8
 Sugawara central charge, 138, 145, 440, 612
 Sugawara construction, 138, 140–142, 145, 185, 368
 Sugawara energy-momentum, 141
 Super-Born-Infeld action, 669
 Superconformal algebra, 149
 Superconformal structure, 556
 Supercurrent, 46
 Superderivative, 557
 Supergravity, 483
 Supermanifold, 556
 (super)moduli space, 265, 556
 Super-Riemann surface, 556, 557
 Superstring, 44
 Supersymmetry, 45
 Supertrace, 573
 Surface group, 43
 Symbol (of an operator), 802
 Symplectic lattice, 311
 Symplectic quotient, 715
 Symplectic representation (of MCG), 312

T

Tadpole, 131, 245, 249, 250, 295, 302, 303, 306, 307, 309, 326, 328, 341–347, 349, 413, 562, 572, 661, 783
 Tate algorithm, 794
 Taub-NUT, 769, 771
 Taub-NUT metric, 770
 T -duality, 376, 462
 Tensor branch, 806, 808
 TFT, 68
 Theory of everything, 5
 θ -function, 331, 400, 545, 684
 3-momentum map, 776
 Topological field theory, 68
 Topological string theory, 430
 Topological twisting, 611
 Topology of surfaces, 26
 Torelli group, 312
 Torsion, 403
 Trace operator, 31
 Triality, 170, 172, 368
 Tube domain, 599
 Twisted chiral superfields, 600
 Twisted sector, 400
 Twisted states, 396
 Twisting, 399

Type IIA, 322

Type IIB, 322

U

Ultraspherical polynomials, 286

Uniformization, 43

V

Vacuum, 78

Variations of Hodge structure, 361

Verma module, 97

Vertex algebra, 320

Vertices, 62

Virasoro algebra, 91

Virasoro central charge, 91

Virasoro constraint, 11

W

Ward identity, 71

Warped product, 510

Warp factor, 510

Wave equation, 11

Weierstrass model, 792

Weil–Petersson metric, 41, 596

Weinberg’s S -matrix theorems, 5

Weinberg–Witten theorem, 5, 491

Wess–Zumino functional, 182

Wess–Zumino–Witten model, 185

Weyl anomaly, 31

Weyl rescaling, 9, 27

Weyl symmetry, 11

Wilson lines, 404

Winding number, 359

Witten index, 156, 337, 490, 663

World-sheet, 4, 8, 9

World-story, 4

World-volume, 411

Y

Yang–Mills, 289

Yau theorem, 592

Yukawa coupling, 6, 642–644, 784

Z

Z^α_β functions, 329, 332

ζ -function, 455

ζ -regularization, 19